YAŞAM 3.0

YAPAY ZEKÂ ÇAĞINDA İNSAN OLMAK

MAX TEGMARK



PEGASUS

"Yapay Zekâ'nın ortaya çıkarabileceği eşi benzeri görülmemiş güç, önümüzdeki yılların insanlığın zirvesi veya çöküşü olabileceği anlamına gelir. Tegmark, Yapay Zekâ'nın sonuçları üzerine bugüne kadar okuduğum en kapsamlı ve eğlenceli kitabı yazmış. Şimdiye dek Tegmark'ın neşeli zekâsıyla tanışmadıysanız sizi çok güzel bir sürpriz bekliyor."
Prof. Erik Brynjolfsson,
MIT Dijital Ekonomi İnisiyatifi Direktörü

"Yapay Zekâ, zekâ ve insanlığın geleceği üzerine düşünüş şeklimizi değiştirecek, harekete geçirici bir kitap."
Bert Selman, Cornell Üniversitesi
Bilgisayar Bilimleri Profesörü

"Tegmark, bir tür olarak nasıl bir gelecek yaratmak istediğimiz hakkında çok daha geniş bir tartışma yaratmak istiyor. Her ne kadar, Yapay Zekâ, kozmoloji, değerler ve hatta bilinçli deneyimin doğası gibi işlediği konular zorlayıcı olsa da bunları okunaklı bir biçimde sunarak okurun kendi fikirlerini oluşturmasına yardım ediyor."
Nick Bostrom, İnsanlığın Geleceği
Enstitüsü'nün Kurucusu

"Tegmark'ın yeni kitabı çağımızın en önemli konusuna dair derinlemesine ve düşündürücü bir rehber. Biyolojik düşüncelerimizi, kendi yarattığımız çok daha büyük bir zekâyla birleştirirken gelecekte nasıl daha iyi bir medeniyet yaratabileceğimizi araştırıyor."
Ray Kurzweil

"Bu kitap beni olduğum yere çiviledi. Yapay Zekâ'nın dönüştürücü sonuçlarıyla kısa süre sonra yüzleşebiliriz ancak bunlar bizim için cennet mi yoksa cehennem mi olacak? Buna karar verebilecek bir jüri henüz yok ama bu aydınlatıcı ve okunaklı kitap bir fikir edinmemizi sağlıyor."
Prof. Martin Rees, Kraliyet Astronomu

"Ünlü bir fizikçi ve Yaşamın Geleceği Enstitüsü başkanı olmak, Max Tegmark'a çağımızın en önemli konusu hakkında okuyucusunu hafife almadan bilgilendirme olanağını sağlıyor."
Jaan Tallinn, Skype'ın Eş Kurucusu

YAŞAM 3.0

Pegasus Yayınları: 2029
Popüler Bilim: 4

Yaşam 3.0
Max Tegmark
Özgün Adı: Life 3.0: Being Human in the Age of Artificial Intelligence

Yayın Koordinatörü: Yusuf Tan
Editör: Kemal Küçükgedik
Düzelti: Halûk Kürşad Kopuzlu
Sayfa Tasarımı: Meral Gök

Baskı-Cilt: Alioğlu Matbaacılık
Sertifika No: 11946
Orta Mah. Fatin Rüştü Sok. No: 1/3-A
Bayrampaşa/İstanbul
Tel: 0212 612 95 59

1. Baskı: İstanbul, Şubat 2019
ISBN: 978-605-299-613-3

Yayıncı Sertifika No: 12177

Pegasus Yayıncılık Tic. San. Ltd. Şti.
Gümüşsuyu Mah. Osmanlı Sk. Alara Han
No: 11/9 Taksim / İSTANBUL
Tel: 0212 244 23 50 (pbx) Faks: 0212 244 23 46
www.pegasusyayinlari.com / info@pegasusyayinlari.com

 pegasusyayinlari  pegasusyayinevi  pegasusyayinlari  Pegasus Yayınları

Max Tegmark

YAŞAM 3.0



Yapay Zekâ Çağında İnsan Olmak

İngilizceden çeviren:

Ekin Can Göksoy

PEGASUS YAYINLARI

Her şeyi mümkün kılan FLI ekibine

İçindekiler

Teşekkür

Bu kitabı yazmaya beni teşvik eden ve bana yardım eden herkese teşekkürlerimi sunarım.

Aileme, arkadaşlarıma, hocalarıma, meslektaşlarıma ve ortaklarıma yıllar süren destek ve ilhamları için.

Anneme bilinç ve anlam üzerine merakımı alevlendirdiği için.

Babama dünyayı daha iyi bir yer yapmak içini mücadele eden ruhu için.

Oğullarım Philip ve Alexander'a, ortaya çıkan insan seviyesi zekânın mucizelerini gösterdikleri için.

Yıllar boyunca benimle iletişime geçen, soruları, yorumları ve fikirlerimi takip edip yayımlatmaktaki destekleri için dünyanın dört bir yanındaki tüm bilim ve teknoloji tutkunlarına.

Menajerim John Brockman'a bu kitabı yazmaya ikna olana kadar beni zorladığı için.

Bob Penna, Jesse Thaler ve Jeremy England'a sırasıyla, kuasarlar, sfaleronlar ve termodinamik hakkındaki faydalı konuşmaları için.

Aşağıda ismi geçen herkese metnin çeşitli kısımlarına dair fikirler ve yorumlar verdikleri için: annem, kardeşim Per, Luisa Bahet, Rob Bensinger, Katerina Bergström, Erik Brynjolfsson, Daniela Chita, David Chalmers, Nima Deghani, Henry Lin, Elin Malmsköld, Toby Ord, Jeremy Owen, Lucas Perry, Anthony Romero, Nate Soares ve Jaan Tallinn.

Kitabın taslaklarına yorumlar yapan şu süper kahramanlara: Meia, babam, Anthony Aguirre, Paul Almond, Matthew Graves,

Phillip Helbig, Richard Mallah, David Marble, Howard Messing, Luiño Seoane, Marin Soljačić, editörüm Dan Frank ve en çok da sevgili ilham perim ve yoldaşım Meia'ya sonsuz teşviği, desteği ve ilhamı için. O olmasaydı bu kitap var olamazdı.

YAŞAM 3.0

İnsanüstü Yapay Zekânın (YZ) bu yüzyılda ortaya çıkacağına inanıyor musunuz?

Evet

Hayır

Bölüm 1'e (sayfa 39) atlayın.

Bir sonraki sayfadan devam edin.

Giriş:



Omega Ekibinin Hikâyesi

Omega ekibi şirketin ruhuydu. İşletmenin kalan kısmı basit YZ'nin çeşitli ticari uygulamaları ile işleri sürdürecek parayı sağlarken, Omega ekibi CEO'nun hep hayal ettiği şeye doğru ilerlemeyi sürdürüyordu: genel yapay zekâyı inşa etmek. Çalışanların büyük bir kısmı, onlara sevgiyle verdikleri isimle, "Omegalar'ı" hedeflerinden daima onlarca yıl geride olan bir grup naif hayalci olarak görüyordu. Yine de onları memnuniyetle şımartıyorlardı çünkü Omegalar'ın şirketlerine kattıkları son teknoloji işin prestiji hoşlarına gidiyordu. Ayrıca Omegalar'ın devamlı olarak sağladıkları iyileştirilmiş algoritmalardan da memnundular.

Fakat Omegalar'ın bu imajı bir sırrı saklamak için özenle inşa etmiş olduğundan habersizdiler: İnsanlık tarihinin en cüretkâr planını gerçekleştirmeye çok yakındılar. Karizmatik CEO'ları onları özenle seçerken yalnızca mükemmel araştırmacılar olmalarını değil, hırslarını, idealizmlerini ve insanlığa yardımcı olmaya kendilerini adamış olmalarını da göz önüne almıştı. Onlara planlarının son derece tehlikeli olduğunu ve güçlü hükümetlerin bunu öğrenmesi durumunda projeyi sonlandırmak ya da tercihen kodlarını çalmak için mümkün olan her şeyi –onları kaçırmak da dâhil– yapabileceklerini hatırlatmıştı. Dünyanın en iyi fizikçilerinin büyük bir kısmının nükleer silahları geliştirmek için Manhat-

tan Projesi'ne dâhil olmalarıyla aynı sebepten ötürü onlar da bu projeye her şeylerini adamışlardı: Eğer bunu ilk olarak kendileri yapmazsa, daha az idealist birilerinin yapacağından emindiler.

İnşa ettikleri Prometheus takma isimli YZ giderek daha da kabiliyetli oluyordu. Bilişsel yetileri halen daha sosyal kabiliyetler gibi pek çok alanda insanlardan oldukça geride olsa da Omegalar bir tek görevde mükemmelleşmesi için ellerinden geleni yapıyorlardı: YZ sistemleri programlamakta. Bu stratejiyi bilerek seçmişlerdi çünkü 1965'te Britanyalı matematikçi Irving Good'un sunduğu zekâ patlaması argümanına inanıyorlardı: "İleri zekâlı bir makineyi en zeki insanın tüm entelektüel aktivitelerini kolaylıkla aşabilen bir makine olarak tanımlayalım. Makine tasarımı bu entelektüel aktivitelerden biri olduğu için, ileri zekâlı makine insanlardan daha iyi makineler tasarlayabilir; bu durumda da tartışmasız biçimde bir 'zekâ patlaması' ortaya çıkacak ve insan zekâsı fersah fersah geride bırakılacaktır. Bu sebeple de makinenin bize kendini nasıl kontrol altında tutacağımızı söyleyecek kadar uysal olması kaydıyla, ilk ileri zekâlı makine insanın yapması gereken son icat olacaktır."

Eğer bu tekrarlamalı öz gelişimi başlatabilirlerse makinenin kısa sürede, faydalı olabilecek tüm diğer insan yetilerini kendine öğretecek kadar zeki olacağını keşfettiler.

İlk Milyonlar

Süreci başlatınaya karar verdiklerinde bir cuma sabahı saat dokuzdu. Prometheus geniş, erişimi kontrol altında tutulan, klimalı bir odada uzun raf sıraları üzerinde duran özel tasarlanmış bilgisayar kümesi içerisinde harıl harıl çalışıyordu. Güvenlik sebeplerinden dolayı, internete hiçbir şekilde bağlı değildi ancak eğitim verisi olarak kullanmak için internetin büyük bir kısmının (Vikipedi,

Kongre Kütüphanesi, Twitter, YouTube'un bir kısmı, Facebook'un büyük bir kısmı vb.) yerel bir kopyasını içeriyordu.* Bu başlangıç zamanını rahatsız edilmeden çalışmak için seçmişlerdi: Aileleri ve arkadaşları hafta sonu bir şirket tatilinde olduklarını düşüneceklerdi. Küçük mutfak mikrodalgaya atılabilir yemek ve enerji içecekleriyle doluydu, başlamaya hazırdılar.

Sistemi açtıklarında, Prometheus YZ sistemleri programlamada onlardan birazcık daha kötüydü ancak bunu, önemli ölçüde hızlı olarak ve onların bir Red Bull içmekte harcadıkları zamanda binlerce kişi-yıl eşdeğerinde iş yaparak telafi ediyordu. Sabah saat on sularında, kendinin bir yeniden tasarımını, v2.0, bitirmişti. Bu görece daha iyiydi ama halen daha insan seviyesinin altındaydı. Öğlen saat iki gibi Prometheus 5.0 faaliyete geçtiğinde, Omegalar dehşet içindeydi: Performans ölçütlerini hezimete uğratmıştı ve gelişim hızı giderek artıyordu. Gece olduğunda, Prometheus 10.0'ı planlarının ikinci aşamasını başlatmakta kullanmaya karar verdiler: para kazanmakta.

İlk hedefleri MTurk, Amazon Mechanical Turk idi. 2005'te kitle kaynaklı bir internet pazarı olarak açıldıktan sonra, İZG, "İnsan Zekâ Görevleri" denen karmaşık yapılı rutin işleri gerçekleştirmek için gece ve gündüz anonim biçimde rekabet eden on binlerce insan sayesinde hızla büyümüştü. Bu rutin işler arasında ses kayıtlarının çözümlenmesi ile web sayfalarının görselleri ve yazılı tanımlamalarının sınıflandırılması gibi görevler bulunuyordu ve tek bir ortak noktaları vardı: Eğer bunu iyi bir şekilde yaparsanız, kimse sizin bir YZ olduğunuzu bilemezdi. Prometheus 10.0 bu görev kategorilerinin yaklaşık yarısını kabul edilebilir

* Çoğu araştırmacı insan seviyesi genel YZ'nin en azından onlarca yıl uzakta olduğunu tahmin etse de bu hikâyede kolaylık olsun diye bugünün ekonomi ve teknolojisini kullandım. Eğer dijital ekonomi giderek büyürse ve çok daha fazla hizmet sorgusuz sualsiz çevrimiçi olarak sağlanabilirse Omega planını gerçekleştirmek çok daha kolaylaşır.

şekilde iyi yapmayı başarmıştı. Bu görev kategorilerinin her biri için, Omegalar Prometheus'a tam olarak bu görevleri yapan ve başka hiçbir şey yapmayan, özel olarak şekillendirilmiş basit bir YZ yazılım modülü tasarlatmıştı. Sonra da bu modülü istedikleri kadar çok sanal makinede çalışabilen bir bulut bilişim platformu olan Amazon Web Hizmetleri'ne yüklediler. Amazon'un bulut bilişim bölümüne ödedikleri her bir dolar için, Amazon'un MTurk bölümünden iki dolardan fazla kazanıyorlardı. Böylesine inanılmaz bir arbitraj fırsatının kendi şirketlerinde var olabileceğinden Amazon hiç şüphelenmemişti!

İzlerini gizlemek için tedbiri elden bırakmayarak önceki aylarda hayalî insanların isimleriyle binlerce MTurk hesabı açmışlardı ve Prometheus'un inşa ettiği modüller şu anda onları bekleyen kimlikleriyle buluşuyordu. MTurk müşterileri normalde sekiz saat kadar sonra ödeme yapıyorlardı, bu noktada da Omegalar devamlı gelişen Prometheus'un son versiyonunun yaptığı çok daha iyi görev modüllerini kullanarak parayı daha fazla bulut bilişim zamanına yatırıyorlardı. Paralarını her sekiz saatte bir iki katına çıkarabildikleri için, kısa süre sonra MTurk'ün görev tedarikini doygunluğa ulaştırmaya başladılar ve dikkat çekmeden günde bir milyon dolardan fazla kazanamayacaklarını fark ettiler. Böylece bir sonraki adıma geçmeleri için gereğinden fazla parayı kazanacaklardı; üstelik artık CFO'dan devamlı nakit istemek zorunda da kalmayacaklardı.

Tehlikeli Oyunlar

YZ atılımlarının yanı sıra, Omegalar'ın en çok eğlendikleri projelerinden biri de Prometheus'un başlatılmasından sonra mümkün olan en hızlı sürede nasıl para kazanacaklarını planlamaktı. Aslında tüm dijital ekonomi önlerine serilmişti ancak bilgisayar oyunları, müzik, filmler ve yazılımlar yaparak başlamak, kitap-

lar veya makaleler yazmak, borsada işlem yapmak ya da keşifler yapıp onları satmak daha mı iyiydi? Önemli olan yatırımın getiri oranını azami seviyeye çıkarmaktı ama normal yatırım stratejileri yapabileceklerinin yavaş çekim bir parodisiydi: Normal bir yatırımcı *yıllık* %9'luk bir getiri oranından memnun kalırken, onların MTurk yatırımları *saatte* %9 getiri bırakıyor, her gün sekiz kat daha fazla para üretiyordu. MTurk'ü doygunluğa ulaştırmışlardı, sırada ne vardı?

İlk düşünceleri borsada vurgun vurmaktı – sonuçta, neredeyse hepsi geçmişte serbest yatırım fonları için YZ geliştirme tekliflerini reddetmişlerdi, bu fonlar da ciddi biçimde tam da bu fikre yatırım yapıyordu. Bazıları *Transcendence* filminde YZ'nin ilk milyonlarını böyle yaptığını hatırlıyordu. Ancak geçen yılki çöküşten sonra vadeli işlemlere getirilen yeni düzenlemeler seçeneklerini kısıtlıyordu. Diğer yatırımcılardan çok daha iyi getiri oranlarına ulaşsalar da kendi ürünlerini satmaktan kazanacaklarının yanına bile yaklaşamayacaklarını kısa sürede fark ettiler. Dünyanın ilk süper zeki YZ'si sizin için çalışıyorsa, diğerlerine değil de kendi şirketlerinize yatırım yapmak daha kârlıydı! Bazı istisnalar elbette olabilir (örneğin Prometheus'un insanüstü hack yetileriyle içeriden bilgi alıp yükselmek üzere olan hisselerin alım opsiyonlarını satın almak gibi) fakat Omegalar bunun yaratacağı istenmeyen ilgiye değmeyeceğini düşünüyorlardı.

Odaklarını kendilerinin geliştirip satabileceği ürünlere doğru kaydırdıklarında, bilgisayar oyunları açık bir şekilde ilk tercih olmalı gibi gözüktü. Prometheus herkese hitap eden oyunları tasarlamada, kodlama, grafik tasarım, görsellerin ışın izlemesi ve son ürünü üretmek için gereken tüm diğer görevlerde hızla beceri geliştirebilirdi. Dahası, insanların tercihleri üzerine tüm web verisini hazmettikten sonra, her oyuncu kategorisinin neyi sevdiğini bilebilecek ve satış gelirleri için bir oyun optimize ede-

cek insanüstü yetiyi geliştirebilecekti. Çoğu Omega'nın itiraf etmek istemeyeceği kadar uzun süre oynadığı bir oyun olan *The Elder Scrolls V: Skyrim* 2011 yılındaki ilk haftasında dört yüz milyon dolardan daha fazla hasılat yapmıştı ve Prometheus'un da bir milyon dolarlık bulut bilişim kaynaklarını kullanarak yirmi dört saat içinde en az bu oyun kadar bağımlılık yaratıcı bir şey üretebileceğine emindiler. Böylece oyunu internet üzerinden satabilecekler ve Prometheus'u oyun hakkında konuşan insanları taklit etmesi için kullanabileceklerdi. Eğer bu, haftada iki milyon dolar getirirse, yatırımlarını sekiz günde sekiz kere ikiye katlayabilecekler, bu da onlara saatlik %3'lük bir getiri sağlayacaktı; MTurk başlangıçlarından hâlâ daha kötüydü ama daha sürdürülebilirdi. Her gün bir dizi başka oyunlar da geliştirerek, kısa süre içinde oyun pazarını doygunluğa ulaştırmadan on milyar dolar kazanabileceklerini fark ettiler.

Ancak ekipteki bir siber güvenlik uzmanı bu oyun işine girmekten onları vazgeçirdi. Bu Prometheus'un *özgürlüğünü kazanması* ve kendi kaderinin kontrolünü eline alması gibi kabul edilemez bir risk taşıyordu. Yinelemeli öz gelişim sürecinde hedeflerinin nasıl evrileceğinden emin olmadıkları için, güvenli şekilde ilerlemeye ve Prometheus'u internete kaçamayacağı biçimde kapalı ("hapiste") tutarak devam etmeye karar verdiler. Sunucu odalarında çalışan ana Prometheus makinesi için fiziksel kısıtlama uyguluyorlardı: Hiçbir şekilde internet bağlantısı yoktu ve Prometheus'tan gelen tek çıktı Omegalar'ın kontrol ettiği bir bilgisayara gönderilen mesajlar ve dokümanlar şeklindeydi.

İnternet bağlantılı bir bilgisayarda ise Prometheus'un ürettiği herhangi bir karmaşık programı çalıştırmak riskliydi: Omegalar makinenin ne yapacağını tam olarak anlayamayacakları için, kendisini viral bir şekilde internete saçmayacağından da emin olamıyorlardı. Prometheus'un MTurk görevleri için yazdığı yazı-

lımları test ederken, bu yazılımları yalnızca sanal bir makinenin içinde çalıştırmak suretiyle bu ihtimalden kendilerini korudular. Bu, bilgisayarı taklit eden bir programdı: Mesela, pek çok Mac kullanıcısı Windows programlarını bir Windows makinesinde olduklarına inandırarak çalıştıran sanal makine yazılımları alırlar. Omegalar da Pandora'nın Kutusu isimli kendi sanal makinelerini üretmişlerdi. Bu makine, normalde bilgisayarlarda bulunan hiçbir özelliği ve parçayı içinde bulundurmayan aşırı basitleştirilmiş bir makineydi: klavye yoktu, monitör yoktu, hoparlör yoktu, internet bağlantısı yoktu, hiçbir şey yoktu. MTurk ses deşifreleri için, Omegalar her şeyi Pandora'nın Kutusu'na giden tüm veri tek bir ses dosyası, çıktı ise tek bir yazı dosyası –deşifraj– olacak şekilde ayarlamıştı. İçindeki yazılımın tabi olduğu kutunun kuralları evrenimizde tabi olduğumuz fizik kuralları gibiydi: Ne kadar zeki olursak olalım ışık hızından hızlı hareket edemeyeceğimiz gibi, yazılım da kutunun dışına çıkamıyordu. O tekil girdi ve çıktı dışında, Pandora'nın Kutusu'ndaki yazılım kendi hesap kuralları içerisinde bir paralel evrende etkili bir şekilde kapana kısılıydı. Omegalar'da öylesine güçlü bir kaçma paranoyası vardı ki zaman kısıtlamasını da devreye sokarak, güvenilir olmayan bir kodun ömrünü sınırladılar. Örneğin, hapsedilmiş deşifre yazılımı her bir ses dosyasını deşifre ettiğinde Pandora'nın Kutusu'nun tüm bellek içeriği otomatik olarak siliniyordu ve program baştan kuruluyordu. Bu şekilde, bir sonraki deşifre görevine başladığında, daha önce ne olduğuna dair hiçbir bilgisi olmuyor, böylece zamanla öğrenme şansı kalmıyordu.

Omegalar MTurk projeleri için Amazon bulutunu kullandıklarında, Prometheus'un oluşturduğu görev modüllerini buluttaki sanal kutucuklara koymayı başarabilmişlerdi çünkü MTurk girdi ve çıktıları çok basitti. Ancak bu, grafik yönünden yoğun olan bilgisayar oyunları için geçerli olmazdı çünkü oyuncunun bilgi-

sayarının tüm donanımına bütüncül bir erişime ihtiyaçları vardı. Dahası, bilgisayar meraklısı bir kullanıcının oyun kodlarını analiz edip Pandora'nın Kutusu'nu keşfetmesini ve içinde ne olduğunu araştırmaya karar vermesini istemezlerdi. Kaçma riski yüzünden yalnızca oyun pazarı elenmemişti, aynı zamanda hazır bekleyen yüzlerce milyar doların olduğu ileri seviye kazançlı diğer yazılım pazarlarının da üstü çizilmişti.

İlk Milyarlar

Omegalar araştırmalarını yüksek katma değerli, tamamen dijital (yavaş üretimi engellemek için) ve kolay anlaşılabilir (mesela, kaçış riski taşımayacağını bildikleri metinler ya da filmler) ürünlere odaklamışlardı. Sonunda, bir medya şirketi kurmaya karar verdiler ve animasyonlarla başladılar. Web sitesi, pazarlama planı ve basın bültenleri Prometheus süper zeki olmadan önce bile hazırdı – tek eksik içerikti.

MTurk'ten düzenli bir şekilde para kıran Prometheus, pazar sabahı olduğunda şaşırtıcı bir biçimde yeteneklerini göstermiş olsa da entelektüel yetileri hâlâ kısıtlıydı: Prometheus daha ziyade sıkıcı MTurk görevlerini yerine getirecek YZ sistemleri tasarlayacak ve yazılımlar yapacak şekilde optimize edilmişti. Örneğin film yapmada kötüydü. Özellikle belirli bir sebepten ötürü kötü değildi, nasıl ki James Cameron doğduğunda film yapmakta kötüyse Prometheus da öyleydi: Bu öğrenmesi zaman alan bir yetiydi. Bir insan çocuğu gibi Prometheus da erişimi olan veriden ne isterse öğrenebilirdi. James Cameron'un okuma yazmayı öğrenmesi yıllar alırken, Prometheus bunu cuma günü, tüm Vikipedi'yi ve birkaç milyon kitabı okumaya fırsat bulduğunda halletmişti. Film yapmak ise daha zordu. İnsanların ilginç bulacağı bir senaryo yazmak, kitap yazmak kadar güçtü. İnsan

toplumuna ve insanların neyi eğlenceli bulduğuna dair detaylı bir anlayışa ihtiyaç vardı. Senaryoyu final video dosyasına dönüştürmek de simülasyon oyuncuların ve içinde yer aldıkları karmaşık sahnelerin yüksek miktarda ışın izlemesi ve simülasyon seslerin, zorlu müziklerin üretimi demekti. Pazar sabahı, Prometheus iki saatlik bir filmi bir dakikada izleyebiliyordu. Bu bir dakikaya filmin uyarlandığı kitabı ve filmin tüm eleştirilerini okumak da dâhildi. Omegalar, birkaç yüz filmi ardı arkasına izleyince Prometheus'un bir filmin ne gibi eleştiriler alabileceğini ve farklı kitlelere nasıl hitap edebileceğini tahmin etmekte iyileştiğini fark ettiler. Gerçekten de sanki gerçek içgörülere yer veriyormuş, senaryodan oyunculuğa, ışık ve kamera açısı gibi teknik detaylara kadar birçok konuda yorum yapıyormuş etkisi veren kendi film eleştirilerini kaleme almaya başladı. Bunun, Prometheus kendi filmlerini yaptığında nasıl başarılı olacağını bileceği anlamına geldiğini düşündüler.

Omegalar, simülasyon oyuncuların kim olduğu konusunda utandırıcı sorulardan kaçınmak için Prometheus'a önceliği animasyona vermesini söylediler. Pazar gecesi, bu çılgın hafta sonunu bira ve mikrodalga fırında yapılmış patlamış mısırla taçlandırarak ışıkları kapadılar ve Prometheus'un ilk filmini izlediler. Disney'in *Frozen* filmine benzer bir fantezi-komedi animasyonuydu ve ışın izleme, bir milyon dolarlık günlük MTurk kârının büyük bir kısmını kullanarak Amazon bulutunda hapsedilmiş Prometheus menşeli kod tarafından yapılmıştı. Film başladığında, filmin insan yönlendirmesi olmadan bir makine tarafından yapılmış olmasını hem büyüleyici hem de korkutucu buldular. Kısa bir süre içinde ise şakalara gülüyorlar, dramatik anlarda nefeslerini tutuyorlardı. Hatta duygusal sona geldiklerinde bazıları biraz ağladı bile, bu kurgusal gerçekliğe öylesine kaptırmışlardı ki kendilerini, yaratıcılarını unutuvermişlerdi.

Omegalar web sitesinin açılışını cuma gününe koyarak Prometheus'a biraz daha içerik üretmek, kendilerine de Prometheus'a güvenmedikleri işleri yapmak için zaman tanımışlardı: Yani reklam almak ve önceki aylarda kurdukları paravan şirketler için çalışan almaya başlamak. İzlerini gizlemek için resmî anlatıları medya şirketlerinin (ki bu şirketlerle Omegalar'ın hiçbir umumi bağlantısı yoktu) içeriğin büyük bir kısmını bağımsız film yapımcılarından, genel olarak da düşük gelir bölgelerindeki yüksek teknoloji startuplarından aldığı yönünde olacaktı. Bu sahte tedarikçiler meraklı gazetecilerin bile gitmeye yanaşmayacağı Tiruchirappalli ve Yakutsk gibi uzak bölgelerde yer alacaklardı. Orada çalışması için işe alınmış elemanlar da pazarlama ve yönetimde çalışacak ve soranlara üretim ekibinin farklı bir konumda olduğunu ve röportaj vermediğini söyleyecekti. Bu hikâyeye uyacak şekilde şirket sloganlarını "Dünyanın yaratıcı yeteneğini yönlendiriyoruz" olarak seçtiler ve şirketlerini özellikle gelişen dünyadaki yaratıcı insanları son teknolojiyle güçlendirerek fark yaratan bir kurum olarak markalandırdılar.

Cuma olduğunda ve meraklı ziyaretçiler gelmeye başladığında, Netflix ve Hulu gibi çevrimiçi eğlence hizmetlerine benzeyen bir şeyle karşılaştılar ancak daha enteresan farkları vardı. Animasyon dizilerini hiç duymamışlardı. Ama oldukça çekiciydiler; çoğu diziler güçlü bir hikâyesi olan kırk beş dakikalık bölümlerden oluşuyordu, her biri de bir sonraki bölümde ne olacağına dair sizi merakta bırakacak şekilde bitiyordu. Hem de rakiplerinden daha ucuzdular. Her dizinin ilk bölümü ücretsizdi ve diğerlerinin de her birini kırk dokuz sente izleyebiliyordunuz, tüm dizi içinse indirim yapılıyordu. Başta yalnızca üç bölümü olan üç dizi vardı ama yeni bölümler günlük olarak çıkıyordu ve farklı demografilerin ihtiyaçlarını karşılayacak yeni diziler ekleniyordu. Prometheus'un ilk iki haftasında, film yapma becerileri hızla

iyileşti. Yalnızca film kalitesi açısından da değil, artık karakter simülasyonu ve ışın izleme algoritmaları da daha iyiydi. Bu da her bölümün bulut bilişim maliyetini önemli ölçüde azaltıyordu. Sonuç olarak, Omegalar ilk ay düzinelerce yeni dizi çıkarabildiler. Bebeklerden yetişkinlere farklı yaş gruplarını hedef aldılar ve ana dünya dilleri pazarlarına genişleyerek tüm rakiplerine göre önemli ölçüde uluslararası bir site oluşturabildiler. Bazı yorumcular yalnızca sesin değil videoların da çok dilli olmasını şaşırtıcı bulmuşlardı: Mesela, ne zaman bir karakter İtalyanca konuşsa, ağız hareketleri de İtalyan kelimelerini takip ediyordu, hem de karakteristik İtalyan el hareketleriyle birlikte. Prometheus şu anda insanlardan ayırt edilemeyecek simülasyon aktörlerle film yapabiliyor olsa da Omegalar niyetlerini belli etmemek için bundan kaçındılar. Ancak gerçek TV dizileri ve filmlerle rekabet edecek türlerde, yarı gerçekçi animasyon insan karakterleri içeren birçok dizi yayımladılar.

Yayın ağları oldukça bağımlılık yaratıcı oldu ve ciddi bir izleyici artışı yaşadı. Pek çok hayran karakterleri ve senaryoları Hollywood'un beyaz perde yapımlarından bile daha zekice ve ilgi çekici buluyordu. Hatta daha ucuza izliyor olmaktan dolayı çok hoşnuttular. Saldırgan bir reklam politikasıyla desteklenen (yaklaşık sıfıra yakın üretim maliyeti olduğu için Omegalar bunu karşılayabiliyordu), mükemmel medya ilgisi ve kulaktan kulağa yayılan reklamla küresel kârı birkaç ay içinde günde on milyon dolara ulaştı. İki ay içinde Netflix'i solladılar ve üç ay sonrasında günde yüz milyon dolarla Time Warner, Disney, Comcast ve Fox'a rakip olarak dünyanın en büyük medya imparatorlukları arasında yerlerini aldılar.

Bu sansasyonel başarı güçlü bir YZ'ye sahip olduklarına dair spekülasyonlar da dâhil olmak üzere pek çok istenmeyen ilgi çekti. Ancak kârlarının sadece küçük bir kısmını kullandıkları

için Omegalar oldukça güçlü bir dezenformasyon kampanyası başlatabildiler. Gösterişli bir Manhattan ofisinden, yeni işe aldıkları sözcüleri paravan hikâyelerini anlatıyordu. Göstermelik olarak pek çok insan işe alındı. Mesela, dünyanın dört bir yanından yeni diziler geliştirmeleri için senaryo yazarlarıyla anlaşıldı ama hiçbiri Prometheus'u bilmiyordu. Bu karmaşık uluslararası taşeron ağı, çalışanlarının işin büyük kısmının başka yerlerde yapıldığını düşünmelerini sağladı.

Kendilerini korumak ve yüksek bulut bilişimin bazılarını uyandırmasını engellemek için, dünyanın dört bir yanında görünüşte alakasız paravan şirketlerinin sahip olduğu büyük bilgisayar tesisleri kurmak için mühendislerle anlaştılar. Çoğunlukla güneş enerjisiyle çalıştıkları için tesislerin kurulduğu yerlerde yaşayanlara bu tesisler "yeşil veri merkezleri" olarak pazarlanmıştı ama aslında depolamadan çok hesaplamaya odaklıydılar. Prometheus yalnızca kullanıma hazır donanımı kullanarak ve onları üretim zamanını en aza indirecek şekilde hazırlayarak en ince ayrıntısına kadar şablon planlarını tasarlamıştı. Bu merkezleri inşa edip çalıştıranlar içeride neyin hesaplandığını bilmiyorlardı. Bildikleri tek şey Amazon, Google ve Microsoft'unkilere benzer ticari bulut bilişim tesislerini işlettikleri ve tüm satışların uzaktan yönetildiğiydi.

Yeni Teknolojiler

Birkaç aylık bir sürede, Omegalar'ın yönettiği bu iş imparatorluğu Prometheus'un yaptığı insanüstü planlama sayesinde dünya ekonomisinde sürekli daha fazla alana açılmaya başlamıştı. Dünyadaki veriyi dikkatli biçimde inceleyerek, Omegalar'a daha ilk haftasında detaylı bir adım adım büyüme planı sunmuş, veri ve bilgisayar kaynakları geliştikçe bu planı iyileştirmeyi ve incelikli hale getirmeyi sürdürmüştü. Prometheus her şeyi bilen olmaktan

çok uzakta olsa da yetileri bir insanınkinden o kadar ilerideydi ki Omegalar onu görev bilinciyle sorularının hepsine muhteşem cevaplar ve tavsiyeler veren mükemmel bir kâhin olarak görüyorlardı.

Prometheus'un yazılımı, üzerinde çalıştığı insan üretimi görece ortalama donanımdan en iyi şekilde yararlanacak biçimde yüksek oranda optimize olmuştu ve Omegalar'ın tahmin ettiği üzere, Prometheus bu donanımı ciddi biçimde iyileştirmenin yollarını bulmuştu. Kaçmasından korkarak Prometheus'un doğrudan kontrol edebileceği robotik üretim tesisleri kurmayı reddettiler. Bunun yerine, farklı konumlarda dünyanın en iyi bilim insanları ve mühendislerinden yüksek sayılarda işe alım yaptılar ve onlara Prometheus'un yazdığı dâhilî araştırma raporlarını vererek bunların diğer merkezlerdeki araştırmacılardan geldiğini söylediler. Bu raporlar mühendislerin test ettiği, anladığı ve ustalaştığı yeni fiziksel etki ve üretim tekniklerini ayrıntılı şekilde aktarıyordu. Normal insan araştırma ve geliştirme (Ar-Ge) süreçleri elbette ki yıllar sürerdi. Bunun en büyük sebebi yavaş ilerleyen deneme yanılma evreleriydi. Şu anki durum ise çok farklıydı: Prometheus çoktan ileri adımları çözmüştü, böylece sınırlayıcı faktör yalnızca insanların doğru şeyi anlayıp inşa etmeye ne kadar hızlı yönlendirilebileceği olacaktı. İyi bir öğretmen, öğrencilerine bilimi kendi kendine öğreneceklerinden daha hızlı öğrenmelerinde yardımcı olur. Prometheus da gizliden gizliye aynı şeyi bu araştırmacılara yapıyordu. Prometheus çeşitli araçlar verildiğinde insanların bir şeyi öğrenip inşa etmesinin ne kadar süreceğini doğru bir şekilde tahmin edebildiği için, ilerlenebilecek en hızlı yolu geliştiriyor, daha kolay anlaşılabilecek, inşa edilebilecek ve daha iyi araçların geliştirilmesinde faydalı olabilecek yeni araçlara öncelik tanıyordu.

Maker kültürüyle, mühendislik ekipleri kendi makinelerini kullanarak daha iyi makineler inşa etmeleri için teşvik edildi. Bu

kendi kendine yeterlilik yalnızca para kazandırmakla kalmadı, ayrıca gelecekte dış dünyadan gelebilecek tehditlere karşı da onları daha korunaklı hale getirdi. İki yıl içerisinde, tüm dünyada üretilenlerden çok daha iyi bilgisayar donanımları üretiyorlardı. Haricî rakiplerine yardım etmemek için, bu teknolojiyi gizlilik içinde tuttular ve yalnızca Prometheus'u geliştirmek için kullandılar.

Dünyanın tanıklık ettiği şey ise şaşırtıcı bir teknolojik patlamaydı. Dünyanın dört bir yanından yeni zengin firmalar neredeyse her alanda devrimci yeni ürünler piyasaya sürüyordu. Güney Koreli bir startup firması laptop pilinizin yarısı büyüklüğünde olmasına ve bir dakikadan az sürede şarj olmasına rağmen iki katı enerji depolayabilen yeni bir pil geliştirmişti. Finlandiyalı bir firma en yakın rakiplerinin iki katı verimliliğe sahip, ucuz bir güneş enerjisi panelini piyasaya sürdü. Bir Alman firması oda sıcaklığında süper iletken özellikleri gösteren ve kitlesel olarak üretilebilen yeni bir tür kablo geliştirerek enerji sektöründe devrim yarattı. Boston merkezli bir biyoteknoloji ekibi yan etkisi olmayan, etkili kilo kaybı ilacının klinik testlerde ikinci aşamaya geçtiğini duyurdu fakat aynı anda Hintli bir ekibin benzeri bir şeyi karaborsada sattığına dair dedikodular da vardı. Californialı bir şirket ise en sık görülen kanserojen mutasyonlara sahip hücreleri tanıyıp saldırabilecek şekilde vücudun bağışıklık sistemini donatabilen, olay yaratacak bir kanser ilacının ikinci aşama testleriyle çıkageldi. Bilimde yeni bir altın çağın başladığına dair tartışmaları tetikleyecek buna benzer birçok örnek ortaya çıkıyordu. Dahası, robotik firmaları dünyanın dört bir yanında mantar gibi bitiyordu. Robotların hiçbiri insan zekâsının yanına bile yaklaşamamıştı ve birçoğu insanlara hiçbir açıdan benzemiyordu. Fakat yine de önemli ölçüde ekonomiyi aksattılar ve takip eden yıllar boyunca, üretim, ulaştırma, depolama, perakende, inşaat, maden, tarım, ormancılık ve balıkçılıkta işçilerin birçoğunun yerini aldılar.

Yetenekli bir grup avukatın çabaları sayesinde dünyanın fark etmediği şey, bu firmaların her birinin bir dizi aracı yoluyla Omegalar tarafından kontrol edildiğiydi. Prometheus dünyanın patent ofislerini farklı temsilciler yoluyla olağanüstü keşiflerle dolduruyordu ve bu buluşlar teknolojinin tüm alanlarında hüküm kurmasına yol açıyordu.

Bu yıkıcı yeni firmalar rekabet içinde güçlü düşmanlar kazanmış olsalar da daha güçlü dostlar da edinmişlerdi. Olağanüstü biçimde kârlıydılar ve "Toplumumuza yatırım yapıyoruz" gibi sloganlarla kârlarının önemli bir kısmını toplum projeleri için tuttukları çalışanlara ödüyorlardı. Bu insanlar da genelde, batmış firmalardan atılmış kişiler oluyordu. Yerel ihtiyaçlara uygun olacak şekilde, en düşük maliyetle hem toplum hem de çalışanlar için maksimum faydayı sağlayacak işleri tanımlamakta Prometheus'un ürettiği detaylı analizleri kullanıyorlardı. Yüksek oranda devlet hizmeti olan bölgelerde, birliktelik oluşturma, kültür, özel ilgi gibi hizmetlere odaklanırlarken, daha fakir bölgelerdeki çalışmaları okul açma ve yönetme, sağlık hizmetleri, çocuk ve yaşlı bakımı, bütçeye uygun konaklama, parklar ve temel altyapı sorunlarını da kapsayabiliyordu. Neredeyse her yerde, yerel sakinler bu hizmetlerin çok önceden yapılması gereken şeyler olduğu konusunda hemfikirdiler. Yerel politikacılar da cömert bağışlar alıyor ve bu kurumsal toplumsal yatırımları teşvik ettikleri için iyi gözükmelerini sağlamak amacıyla özel çaba sarf ediliyordu.

Güç Kazanmak

Omegalar'ın bir medya şirketi kurmalarının tek sebebi erken dönem teknoloji girişimlerini finanse etmek değildi, ayrıca cesur planlarının bir sonraki adımı için de ona ihtiyaç duyuyorlardı: dünyayı ele geçirmek. İlk piyasaya çıkışlarından sonraki bir yıl

içerisinde, tüm dünyadan ileri derecede iyi haber kanallarını saflarına kattılar. Diğer kanallarının aksine, bunlar bilinçli olarak para kaybedecek şekilde tasarlanmıştı ve kamusal hizmet olarak çalışıyorlardı. Aslında haber kanallarının geliri yoktu: hiçbir şekilde reklam almıyorlardı ve internet bağlantısı olan herhangi biri tarafından ücretsiz biçimde izlenebiliyorlardı. Medya imparatorluklarının geri kalanı öylesine çok nakit kazandırıyordu ki dünya tarihinde herhangi bir gazetecilik girişiminden çok daha fazla kaynağı haber servislerine harcayabilmişlerdi – meyvelerini de topladılar. Yüksek oranda rekabetçi gazeteci ve araştırmacı muhabir maaşları ile agresif bir işe alım politikası izleyerek, olağanüstü yetenekleri ve bulguları ekrana taşıdılar. Yerel yolsuzluktan kalpleri ısıtacak bir olaya kadar haber değeri taşıyan herhangi bir şeyi onlara bildiren herkese ödeme yapan küresel bir internet hizmeti yoluyla, bir hikâyeyi ilk aktaran genelde onlar oluyordu. En azından insanların inandığı buydu: aslında, ilk onlar oluyordu çünkü vatandaş gazetecilere atfedilen hikâyeler aslında gerçek zamanlı internet gözetlemesi yoluyla Prometheus tarafından keşfediliyordu. Tüm bu video haber sitelerinin ses ve görüntü akışları ve basılı makaleleri de oluyordu.

Habercilik stratejilerinin birinci aşaması insanların güvenini kazanmaktı, bunu da büyük başarıyla yaptılar. Daha önce eşi görülmemiş para harcama istekleri, izleyicilerinin ilgisini çekecek skandalları araştırmacı gazetecilerin ifşa edebileceği bölgesel ve yerel özenli gazeteciliği mümkün kılabiliyordu. Bir ülke, politik olarak güçlü biçimde kutuplaştığında ve partizan medyalara alışıldığında, her grup için görünüşte farklı firmalara ait olan bir haber kanalı açıyorlar ve o grubun güvenini kazanıyorlardı. Mümkün olduğunda, aracılarını kullanıp var olan en etkili kanalları alıyor, reklamları kaldırıp kendi içeriklerini devreye sokuyorlardı. Sansür ve politik müdahalenin bu tip çabaları tehdit edebileceği

ülkelerde, ilk başta işin içinde kalabilmek için devletin onlardan istediği her şeyi kendi içlerindeki, "gerçek, yalnızca gerçek ama belki de tüm gerçek değil" sloganıyla yapıyorlardı. Prometheus böyle durumlarda hangi politikacıların iyi gösterilmesi gerektiğini ve hangilerinin (genelde de yozlaşmış yerel politikacıların) ifşa edilebileceğini göstererek mükemmel tavsiyeler veriyordu. Prometheus, ayrıca kimin iplerini elde tutmak gerektiğine, kime rüşvet vermek gerektiğine ve bunu en iyi ne şekilde yapmak gerektiğine dair öneriler de getiriyordu.

Bu strateji tüm dünyada inanılmaz bir başarı yarattı. Omega kontrolündeki kanallar en güvenilir haber kaynakları olmuştu. Kitlelerin kanalları benimsemesine hükümetlerin engel olduğu ülkelerde bile, güvenilirlikleri üzerinden bir ün kazanmışlar ve haberlerinin birçoğu da dedikodu olarak yayılmıştı. Rakip haber müdürleri çaresiz bir savaşın içinde olduklarını hissediyorlardı: Daha iyi kaynakları olan ve ürünlerini bedavaya veren biriyle nasıl rekabet edebilirdin ki? İzleyici sayıları düşerken gün geçtikçe daha çok firma kanallarını satmaya karar veriyordu. Genellikle de Omegalar'ın kontrolü altında olduğu ortaya çıkan bir konsorsiyuma satıyorlardı.

Prometheus'un çalışmaya başlamasından iki yıl kadar sonra, güven kazanma kısmı büyük ölçüde tamamlandığında, Omegalar haber stratejilerinin ikinci aşamasını devreye soktular: ikna. Çok daha öncesinde, kurnaz gözlemciler bu yeni medyanın siyasi bir gündemi olduğuna dair ipuçlarını keşfetmişlerdi: Sanki merkeze doğru, her tür aşırılıktan uzağa hafif bir itiş vardı. Bambaşka gruplara hizmet veren bir sürü kanal, halen daha ABD ve Rusya, Hindistan ve Pakistan, farklı dinler, politik gruplar vesaire arasındaki düşmanlığı yansıtıyordu ancak eleştiri dozu yavaş yavaş azalıyor, genelde de kişileri karalamaya yönelik saldırılar, korku yaratma ya da altı doldurulmamış dedikodulardan ziyade para

ve iktidarı içeren somut haberlere odaklanıyorlardı. İkinci aşama ciddi biçimde başladığında, çatışmaları alevlendiren kişilerin aslında nasıl kişisel kâr amacıyla hareket ettiklerine dair araştırmacı gazetecilikle süslenmiş, geleneksel karşıtlıkların acı yüzüne dair sık yayımlanan dokunaklı hikâyelerle, eski çatışmaları ortadan kaldırmaya yönelik bu itki daha da belirginleşmişti.

Siyasi yorumcular, bölgesel çatışmaları yumuşatmaya paralel olarak küresel tehditleri de azaltmaya yönelik planlı bir itkinin var olduğunu ileri sürüyorlardı. Mesela, nükleer savaşın riskleri birden her yerde konuşulmaya başlanmıştı. Kazaen ya da bilinçli bir şekilde küresel bir nükleer savaşın başladığı senaryoları olan ve kitlesel açlığı, altyapının çökmesini, nükleer kışı içeren distopyayı dramatize eden büyük bütçeli filmler çıkmaya başlamıştı. Yeni belgeseller nükleer kışın her ülkeyi nasıl etkileyeceğini ayrıntılandırıyordu. Nükleersizleşmeyi savunan bilim insanlarına ve politikacılara televizyonda daha çok yer veriliyordu. Özellikle de ne gibi adımların atılabileceğine dair yeni çalışmaların sonuçlarını tartışmak için. Bu çalışmalar da yeni teknoloji şirketlerinden büyük bağışlar almış bilimsel organizasyonlar tarafından gerçekleştiriliyordu. Sonuç olarak, füzelerin tetiklerini uzaklaştıracak ve nükleer cephaneliği azaltacak bir politik ivme gelişti. Yeni enerji altyapılarına yatırım yapmaları için hükümetleri teşvik eden ve yenilenebilir enerji maliyetlerini düşüren ve Prometheus'un mümkün kıldığı yeni teknolojik gelişmelere sık sık değinilerek, medya ilgisini küresel iklim değişikliğine kaydırdı.

Medyayı ele geçirmelerine paralel olarak, Omegalar Prometheus'u eğitimde devrim yapmakta da kullandılar. Bir kişinin bilgi ve yetileri verildiğinde, Prometheus yeni bir konuyu en motive ve hevesli olacak şekilde öğrenmesi için en hızlı yolu belirliyor ve optimize videolar, okuma materyalleri, egzersizler ve diğer öğrenme araçlarını üretiyordu. Omega kontrollü şirketler de nere-

deyse her konuda, hem başlangıç seviyesi hem de dil ve kültürel arka plana göre yüksek derecede özelleşmiş çevrimiçi kursları pazarlıyorlardı. İster okumayı öğrenmek isteyen kırk yaşında biri isterseniz de kanser immünoterapisi hakkında son gelişmeleri araştıran biyoloji doktorasına sahip biri olun, Prometheus sizin için mükemmel dersi sunuyordu. Bu dersler günümüz çevrimiçi derslerine neredeyse hiç benzemiyordu: Prometheus'un film yapma yetilerini kullanarak hazırladığı videolar öğrencileri tamamen bağlıyor, bağlantı kurabilecekleri güçlü metaforlar sunuyor ve daha fazlasını öğrenme isteği yaratıyordu. Bazı dersler ücretli olsa da pek çoğu ücretsizdi. Öğretmenler için bu bulunmaz bir nimetti çünkü derslerinde kullanabiliyorlardı; aslında bir şeyler öğrenmek isteyen herkes için büyük bir nimetti.

Bu eğitim süper güçleri politik amaçlar için de güçlü araçlardı. Her bir videonun bir kişinin görüşlerini değiştirmesini sağlayacak ve tekrar ikna edilecekleri alakalı bir konu üzerine bir video daha izlemeleri için onları motive edecek iç görülere sahip çevrimiçi video "ikna dizileri" üretiyorlardı. Mesela amaç iki ulus arasındaki bir sorunu çözmek olduğunda, iki ülkede de çatışmanın sebebini ve ilerleyişini incelikli biçimde ortaya koyan tarihsel belgeseller birbirinden bağımsız şekilde yayımlanıyordu. Pedagojik haberler kendi taraflarındaki kişilerden hangisinin devam eden çatışmadan fayda sağladığını ve bunu nasıl yaptığını açıklıyordu. Aynı anda, tıpkı geçmişte sempatizan biçimde gösterilen azınlık karakterlerinin insan hakları ve gey hakları hareketlerini desteklemesi gibi, diğer ulustan sevilebilir karakterler eğlence kanallarında belirmeye başlıyordu.

Kısa süre içinde, siyasi yorumcular yedi slogan etrafında toplanan siyasi bir gündem için artan desteği fark etmişlerdi:

1. Demokrasi
2. Vergi indirimi
3. Hükümet sosyal hizmetlerinde kesintiler
4. Askerî harcamalarda kesintiler
5. Serbest ticaret
6. Açık sınırlar
7. Sosyal sorumluluğa sahip şirketler

Daha az açık olan şey ise altındaki amaçtı: Geçmişin tüm güç yapılarını dünyadan silmek. 2-6 arası hedefler devletin gücünü azalttı ve dünyayı demokratikleştirmek Omegalar'ın iş imparatorluğunun politik liderlerin seçiminde daha etkili olmasını sağladı. Sosyal sorumluluğa sahip şirketler de devletin sağladığı (ya da sağlaması gereken) hizmetleri giderek ele geçirerek devlet gücünü iyice zayıflattı. Geleneksel iş dünyası elitleri de zayıflıyordu çünkü serbest pazarda Prometheus'un desteklediği şirketlerle mücadele edemezlerdi. Bu yüzden de dünya ekonomisinde giderek azalan bir yüzdeye sahip oluyorlardı. Siyasi partilerden inanç gruplarına kadar geleneksel kanaat önderleri Omegalar'ın medya imparatorluğuyla mücadele edecek ikna mekanizmalarından yoksundular.

Tüm köklü değişikliklerde olduğu gibi, kazananlar ve kaybedenler vardı. Eğitim, sosyal hizmetler ve altyapı gelişir, çatışmalar biter ve yerel şirketler dünyayı kasıp kavuran ileri teknolojiler piyasaya sürerken çoğu ülkede somut bir iyimserlik havası esmesine rağmen, herkes mutlu değildi. İşlerinden atılmış işçiler toplum projeleri için yeniden işe alınırken, büyük güç ve zenginlik sahibi olanlar küçülüyordu. Bu, medya ve teknoloji sektörlerinde başladı ama neredeyse her alana yayıldı. Dünyadaki çatışmaların azalması savunma bütçelerinin de azalmasına yol açtı, bu da savunma sanayisi şirketlerinin işlerini bozdu. Yeni yeni filiz veren firmalar

genellikle borsada işlem görmezdi çünkü kâr maksimizasyonunu hedefleyen hissedarların, şirketlerin toplumsal projelere yaptıkları yüksek oranda harcamaları engelleyeceği düşünülürdü. Bu yüzden de küresel borsa piyasaları değer kaybederek hem finans baronlarını hem de emeklilik fonlarına güvenen sıradan insanları tehdit altına aldı. Borsada işlem gören şirketlerin azalan kârları yeterince kötü değilmiş gibi, dünyanın dört bir yanından yatırım firmaları rahatsız edici bir trendi fark ettiler: Geçmişte başarılı olan ticaret algoritmaları en basit indeksli yatırım fonlarında bile başarısız oluyordu. Birileri onları kendi oyunlarında yenmiş gibi gözüküyordu.

Kitlesel olarak güçlü insanlar değişim rüzgârlarına karşı koysa da tepkiler etkisizdi, detaylıca planlanmış bir tuzağa düşmüş gibiydiler. Büyük değişiklikler öylesine baş döndürücü bir hızda oluyordu ki takip etmek ve koordine bir cevap geliştirmek mümkün olmuyordu. Dahası, neyi talep etmeleri gerektiği de tamamen belirsizdi. Geleneksel sağ politika, sloganlarının birçoğunun hayata geçtiğini görüyordu ancak vergi indirimi ve gelişmiş iş dünyası çoğunlukla yüksek teknolojiye sahip rakiplerinin işine yarıyordu. Neredeyse tüm geleneksel endüstri bir kurtarma paketi için yalvarıyordu ancak kısıtlı hükümet fonları onları birbirlerine karşı giriştikleri umutsuz kavgada yalnız bırakmıştı. Öte yandan medya onları rekabeti beceremedikleri için devlet kaynaklarının peşinde koşan dinozorlar olarak gösteriyordu. Geleneksel sol politika serbest ticaret ve hükümetin sosyal hizmet indirimlerine karşı çıkıyordu ama askerî harcamalardaki indirimler ve fakirliğin azalmasından mutluydu. Gerçekten de sosyal hizmetlerin iyileşmesinden ötürü otoriteleri sarsılmıştı. Şu anda bu hizmetleri devlet değil idealist firmalar sağlıyordu. Her yerden anketler insanların yaşam kalitelerinin arttığını ve her şeyin genel olarak iyi bir yönde ilerlediğini gösteriyordu.

Bunun basit bir matematiksel açıklaması vardı: Prometheus'tan önce, dünyanın nüfusunun en fakir %50'si küresel gelirin %4'üne sahipti, bu da Omega kontrolündeki şirketlerin kârlarının küçük bir kısmını paylaşarak onların kalplerini (ve oylarını) çalmasına imkân sağlamıştı.

Konsolidasyon

Sonuç olarak, yedi Omega sloganını benimseyen partiler her ülkede açık ara farkla galibiyet kazandı. Dikkatle optimize edilmiş kampanyalarda, kendilerini politik spektrumun merkezinde gösterdiler. Sağı açgözlü, devletten sübvansiyon bekleyen karmaşa çıkaranlar olarak ifşa ederken solu da hantal hükümet ve vergi meraklısı, yenilikleri baltalayanlar olarak gösterdiler. Neredeyse kimsenin anlamadığı ise Prometheus'un aday olarak öne sürülecek optimal kişileri dikkatle seçtiği ve galibiyetini garanti altına almak için o kişileri yönettiğiydi.

Prometheus'tan önce, teknolojik işsizliğe çare olarak vergilerle fonlanacak evrensel vatandaşlık geliri için artan bir destek vardı. Kurumsal sorumluluk projeleri uçuşa geçince bu hareket de patladı çünkü Omega kontrollü iş imparatorluğu aslında aynı şeyi sunuyordu. Sosyal sorumluluk projelerinin koordinasyonunu iyileştirme bahanesiyle uluslararası bir şirket grubu, Humanitarian Alliance isimli bir sivil toplum örgütünü kurdu. Bu örgüt dünyadaki en değerli insani yardım kampanyalarını bulacak ve finanse edecekti. Kısa süre içinde, neredeyse tüm Omega İmparatorluğu bu örgütü destekledi ve beklenmedik ölçüde küresel projeler ortaya koydu. Bu projeler, teknoloji patlamasında dışta kalan bazı ülkelerde bile eğitim, sağlık, refah ve yönetişim konularında ilerleme sağladı. Elbette, Prometheus perde arkasında bu proje planlarını dikkatle yaparak, planları dolar başına po-

zitif etkiye göre sıralıyordu. Nakdi azar azar dağıtmak yerine, temel gelir teklifleri altında, Alliance (artık böyle biliniyordu) desteklediklerini kendi hedeflerine doğru çalışma konusunda teşvik ediyordu. Sonuç olarak, dünya nüfusunun büyük bir kesimi Alliance'a sadık ve minnettar hissetmeye başlamıştı; genelde, kendi hükümetlerinden daha da çok.

Zaman geçtikçe, Alliance giderek daha da dünya hükümeti görevini üstlenmeye başladı, ulusal hükümetlerin güçleri de devamlı olarak azalıyordu. Ulusal bütçeler vergi indirimlerinden ötürü azalırken Alliance'ın bütçesi tüm hükümetlerin bütçesinin toplamına göre bile devasa sayılırdı. Ulusal hükümetlerin tüm geleneksel rolleri gereksiz ve alakasız gözükmeye başlamıştı. Alliance en iyi sosyal hizmet, eğitim ve altyapıyı sağlıyordu. Medya uluslararası çatışmaları sona erdirmişti, askerî harcamalar gereksiz kalmıştı ve artan refah kıt kaynaklar üzerine rekabete dayanan pek çok eski anlaşmazlığın kökünü kurutmuştu. Birkaç diktatör ve başkaları bu yeni dünya düzenine şiddetli biçimde karşı çıkmıştı ve dâhil olmayı reddetmişlerdi ancak dikkatle düzenlenen darbeler ya da kitlesel ayaklanmalarla her biri devrilmişti.

Omegalar dünyadaki hayatın tarihi boyunca yaşanmış en dramatik değişimi tamamlamışlardı. İlk kez, Dünya'da ve kozmosta yaşamı potansiyel olarak milyarlarca yıl sürdürebilecek güçte bir zekâyla gücünü artıran, tekil bir güç gezegenimizi yönetiyordu ama planları tam olarak neydi?

***

Omega ekibinin hikâyesi buydu. Kitabın kalan kısmı başka bir hikâyeyle ilgili – henüz yazılmamış bir hikâyeyle: YZ ile olan kendi geleceğimizin hikâyesi. Nasıl bitmesini isterdiniz? Omega hikâyesine uzaktan da olsa

benzer bir şey olabilir mi ve eğer olabilirse, ister misiniz? İnsanüstü YZ hakkında spekülasyonları bir kenara bırakırsak, hikâyemizin nasıl başlamasını isterdiniz? YZ'nin önümüzdeki on yılda işlere, hukuka ve silahlara etkisinin nasıl olmasını istersiniz? Geleceğe baktığınızda, siz olsanız sonu nasıl yazarsınız? Bu kesinlikle kozmik boyutlarda bir hikâye çünkü anlattığı, evrenimizde hayatın geleceğinden başka bir şey değil. Bu hikâyeyi yazmak da bize düşüyor.

Bölüm 1:



Çağımızın En Önemli Tartışmasına Hoş Geldiniz

Teknoloji daha önce hiç olmadığı kadar gelişmesi ya da kendini yok etmesi için yaşama potansiyel sağlıyor.

Yaşamın Geleceği Enstitüsü

Doğumundan 13.8 milyar yıl sonra evrenimiz uyandı ve kendinin farkına vardı. Küçük mavi bir gezegenden, evrenimizin küçük bilinçli kısımları kozmosa teleskoplarla bakmaya, var olduğunu sandıkları her şeyin devamlı olarak daha büyük bir şeyin parçası olduğunu keşfetmeye başladılar: bir güneş sistemi, bir galaksi ve incelikli bir gruplar, kümeler ve süperkümeler dizisi içinde düzenlenmiş yüz milyarlarca galaksiyle bir evren. Yıldızları izleyen bu bilinçli kimseler pek çok konuda anlaşamasalar da bu galaksilerin güzel ve büyüleyici olduğunda hemfikirler.

Fakat güzellik göreceli olsa da fizik kuralları değildir. Bu yüzden de evrenimiz uyanmadan önce, güzellik yoktu. Bu da kozmik uyanışımızı daha da muhteşem ve kutlamaya değer yapıyor: Evrenimizi öz bilinci olmayan akılsız bir zombiden kendi üzerine düşünmeyi, güzelliği ve umudu –hedeflerin, anlamın ve amacın peşinden koşmayı– bir araya getiren canlı bir ekosisteme dönüştürdü. Evrenimiz uyanmamış olsaydı, bana göre, tamamıyla

anlamsız olacaktı; tam anlamıyla devasa bir yer israfı. Evrenimiz bir sebeple, mesela kozmik bir kıyamet ya da kendi kendine olan bir talihsizlik yüzünden, yeniden süresiz bir uykuya dalarsa, maalesef yine anlamsız olacak.

Öte yandan, her şey daha da iyiye gidebilir. Kozmosumuzda yıldızları izleyenler yalnızca, ya da ilk izleyenler, biz insanlar mıyız bunu bile bilmiyoruz henüz ancak evrenimizin şu ana dek olduğundan çok daha tam biçimde uyanma potansiyeli olduğunu öğrendik. Belki de bu sabah uykudan kalktığınızda yaşadığınız o ilk zayıf kendi bilincine kavuşma aşamasındayız henüz: Gözlerinizi açıp tamamen uyandığınızda gelecek çok daha büyük bir bilincin öncülü gibi. Belki de yaşam, kozmosumuza yayılacak ve milyarlarca hatta trilyonlarca yıl boyunca gelişecek. Belki de bu, yaşam süremiz boyunca küçük gezegenimizde verdiğimiz kararlardan ötürü olacak.

Karmaşıklığın Kısa Tarihi

Peki, bu inanılmaz uyanış nasıl meydana geldi? Münferit bir olay değildi bu ancak evrenimizi daha da karmaşık ve ilgi çekici yapan –ve halen artan bir ivmeyle devam eden– dur durak bilmeyen 13.8 milyar yıllık bir sürecin içinde bir adımdı.

Bir fizikçi olarak, son çeyrek yüzyılın büyük bir kısmını kozmik tarihimizi tespit etmiş olmaya harcamış olmaktan dolayı şanslı hissediyorum. Bu inanılmaz bir keşif yolculuğu oldu. Lisansüstü öğrencisi olduğum günlerde evrenimizin on milyar yıllık mı yoksa yirmi milyar yıllık mı olduğunu tartışırken bugün 13.7 milyar mı yoksa 13.8 milyar mı diye tartışmaya başladık. Bu elbette çok daha iyi teleskoplar, bilgisayarlar ve çok daha iyi bir kavrayışın eseriydi. Biz fizikçiler halen daha Büyük Patlama'ya neyin yol açtığını ya da bunun gerçekten her şeyin başlangıcı

mı yoksa daha erken bir aşamanın devamı mı olduğunu bilmiyoruz. Ancak, yüksek kalite ölçüm bombardımanı sayesinde, Büyük Patlama'dan *bu yana* neler olduğuna dair görece detaylı bir fikrimiz var. Bu yüzden izninizle 13.8 milyar yıllık kozmik tarihimizi size kısaca özetleyeyim.

Başlangıçta ışık vardı. Büyük Patlama'dan sonraki ilk salisede, prensipte teleskoplarımızın gözlemleyebildiği tüm uzay ("gözlemlenebilir evrenimiz" ya da kısaca "evrenimiz") güneşimizin çekirdeğinden çok daha sıcak ve parlaktı ve hızla genişliyordu. Bu şaşırtıcı gözükse de aynı anda sıradandı da. Çünkü evrenimiz yalnızca cansız, yoğun, sıcak ve sıkıcı derecede tek tip temel parçacıklar çorbasından ibaretti. Her şey neredeyse her yerde aynıydı ve tek ilginç yapı çorbayı bazı yerlerde yaklaşık %0,001 daha yoğun yapan çok hafif ve rastgele gibi gözüken ses dalgalarıydı. Bu hafif dalgaların genel olarak kuantum dalgalanmalarından kaynaklandığına inanılır çünkü Heisenberg'in kuantum mekaniğindeki belirsizlik ilkesi hiçbir şeyin tamamen sıkıcı ve tek tip olmasına izin vermez.

Evrenimiz genişleyip soğurken, parçacıklar birleşerek daha karmaşık nesneler oluşturduğu için ilginçleşmeye de başlamıştı. İlk salise boyunca güçlü nükleer kuvvet, kuarkları protonlara (hidrojen çekirdeği) ve nötronlara dönüştürdü; bunların bazıları da birkaç dakika içinde helyum çekirdeğine dönüştü. Dört yüz bin yıl kadar sonra, elektromanyetik güç bu çekirdekleri elektronlarla bir araya getirerek ilk atomları oluşturdu. Evrenimiz genişlemeyi sürdürürken bu atomlar giderek soğuyarak karanlık bir gaza dönüştü. Bu ilk gecenin karanlığı yaklaşık yüz milyon yıl kadar sürdü. Bu uzun gece, yerçekim kuvveti gazdaki dalgalanmaları artırmayı başarıp atomları ilk yıldız ve galaksileri oluşturmaya ittiğinde, yerini kozmik şafağa bıraktı. Bu ilk yıldızlar hidrojeni

karbon, oksijen ve silikon gibi daha ağır atomlar haline gelecek şekilde birleştirerek ısı ve ışık üretti. Bu yıldızlar öldüğünde, yarattıkları atomların birçoğu kozmosta geri dönüşüme uğradı ve ikinci nesil yıldızların etrafındaki gezegenleri oluşturdular.

Bir noktadan sonra, bir grup atom hem kendi varlığını sürdürebilirken hem de kopyalayabilen karmaşık bir örüntüde birleşti. Kısa süre sonra iki kopyası vardı ve sayı devamlı ikiye katlanıyordu. Bir trilyon olması için sadece kırk kere kopyalama yeterliydi, böylece bu ilk kendini kopyalayıcı kısa sürede dikkate alınacak bir güç halini aldı. Yaşam ortaya çıkmıştı.

Yaşamın Üç Aşaması

Yaşamı nasıl tanımlamak gerektiğine dair soruların tartışmalı olduğu ünlüdür. Birbiriyle yarışan sayısız tanım vardır, bazıları hücrelerden oluşmak gibi gayet spesifik koşullar ileri sürer, bu da geleceğin zeki makinelerini ve dünya dışı uygarlıkları şimdiden diskalifiye etmiş olur. Hayatın geleceğine dair düşünüşümüzü şimdiye dek karşılaştığımız türlerle sınırlamak istemediğimiz için, yaşamı çok daha geniş biçimde tanımlayalım, basitçe karmaşıklığını sürdürebilen ve kendini kopyalayabilen bir süreç olarak. Kopyalanan şey madde (atomlardan yapılmış) değildir, kopyalanan atomların nasıl dizileceğini belirten bilgidir (bitlerden yapılmış). Bir bakteri DNA'sının kopyasını yaptığında, hiç yeni atom üretilmez ancak yeni bir dizi atom orijinaliyle aynı örüntüde düzene girer, böylece de bilgi kopyalanmış olur. Diğer bir deyişle yaşamı, bilgisi (yazılımı) hem davranışına hem de donanımının kılavuzlarına karar veren, kendini kopyalayan ve bilgi üreten bir sistem olarak düşünebiliriz.

Evrenimiz gibi, yaşam da daha karmaşık ve ilgi çekici hale geldi* ve şimdi açıklayacağım üzere, yaşam formlarını üç karmaşıklık seviyesinde sınıflandırmayı faydalı buluyorum: Yaşam 1.0, 2.0 ve 3.0. Bu üç seviyeyi Görsel 1.1'de özetledim.

Evrenimizde yaşamın ilk olarak nasıl, ne zaman ve nerede ortaya çıktığı halen cevaplanamamış bir sorudur ancak Dünya'da yaşamın ilk kez dört milyar yıl önce ortaya çıktığına dair güçlü kanıtlar vardır. Kısa süre içinde, gezegenimiz birbirinden farklı her tür yaşam formuyla dolmuştu. Kısa süre içinde diğerlerini alt edecek olan en başarılıları bir şekilde çevrelerine tepki gösterebilmiş olanlardı. Özel olarak, bilgisayar bilimcilerin "akıllı üstlenici" dediği şeylerdi: sensörlerden çevreleri hakkında bilgi toplayan, sonra da bu bilgiyi çevrelerine nasıl karşılık vereceklerine karar vermek için işleyen varlıklar. Buna, göz ve kulaklarınızdan gelen bilgiyi bir konuşmada ne diyeceğinizi söylemeye karar verirken kullanmanız gibi yüksek derecede karmaşık bilgi işlemler de dâhil edilebilir. Ancak oldukça basit yazılım ve donanımı da kapsıyor olabilir.

Mesela, pek çok bakteri etraflarındaki sıvının şeker konsantrasyonunu ölçmeye yarayan bir duyargaya ve kamçı denen pervane şekilli yapılara sahiptir. Duyargayı kamçıya bağlayan donanım şu basit ama işlevsel algoritmayı kullanıyor olabilir: "Eğer şeker konsantrasyonu duyargam birkaç saniye öncesindekine göre daha düşük bir değer rapor ediyorsa, kamçımın yönünü tersine çevir ve böylece yönümü değiştir."

* Yaşam neden daha karmaşık bir hal aldı? Evrim, çevresindeki düzeni tahmin edecek ve kullanacak kadar karmaşık yaşamı ödüllendirir, böylece daha karmaşık bir çevrede, daha karmaşık ve zeki bir yaşam evrimleşir. Bu zeki yaşam da rekabet içindeki yaşam formları için daha karmaşık bir çevre yaratır, karşılığında da bu yaşam formları daha da karmaşık olur ve nihayetinde, ileri derecede kompleks yaşam ekosistemi ortaya çıkar.

Donanımını tasarlayabilir mi?
Yazılımını tasarlayabilir mi?
Hayatta kalabilir ve kendini kopyalayabilir mi?

Hoşça kalın!
¡Hola!
¡Hola!
Hey!
Hey!

Yaşam 1.0 (basit biyolojik)
Yaşam 2.0 (kültürel)
Yaşam 3.0 (teknolojik)

Görsel 1.1: Yaşamın üç aşaması: biyolojik evrim, kültürel evrim ve teknolojik evrim. Yaşam 1.0 yaşamı boyunca ne yazılımını ne de donanımını tasarlayabilir: İkisi de DNA'sı tarafından belirlenir ve yalnızca evrim yoluyla pek çok nesil sonrasında değişebilir. Aksine Yaşam 2.0, yazılımının büyük bir kısmını yeniden tasarlayabilir: İnsanlar karmaşık yeni yetiler kazanırlar –mesela, dil, spor ve meslek– ve dünya görüşleri ile hedeflerini güncelleyebilirler. Henüz Dünya üzerinde yer almayan Yaşam 3.0 yalnızca yazılımını değil, donanımını da nesiller boyunca yavaşça değişmesini beklemeden yeniden tasarlayabilir.

Konuşmayı ve sayısız diğer şeyi öğrendiniz. Bakteriler ise pek iyi öğrenciler değillerdir. DNA'ları yalnızca donanımlarının, mesela şeker duyargası ve kamçı, tasarımını vermekle kalmaz, yazılımlarını da verir. Asla şekere doğru yüzmeyi öğrenmezler; bunun yerine, bu algoritma DNA'larında vardır. Elbette bu da bir nevi eğitim süreci gerektirir ama bu bir bakterinin yaşam döngüsü boyunca gerçekleşmez. Bu, doğal seleksiyonun şeker tüketimini

iyileştiren rastgele DNA mutasyonlarını seçtiği, nesillere yayılan yavaş bir deneme yanılma süreciyle o cins bakterinin önceki evrimi sürecinde gerçekleşmiştir. Bu mutasyonların bir kısmı kamçı ve diğer donanımın tasarımının da iyileşmesine yardımcı olurken, diğer mutasyonlar şeker bulma algoritması ve diğer yazılımlara katkı sunan bakteriyel bilgi işleme sistemini geliştirmiştir.

Bu cins bakteriler "Yaşam 1.0" dediğim türe örnektir: *yazılım ve donanımın tasarlanmış olmaktan çok evrimleşmiş olduğu yaşam.* Siz ve ben ise "Yaşam 2.0" örnekleriyiz: *donanımı evrimleşmiş ama yazılımı büyük ölçüde tasarlanmış yaşam.* Yazılımınız derken, duyularınızdan topladığınız bilgiyi işlemekte ve ne yapacağınıza karar vermekte kullandığınız algoritmalar ve bilgiyi kastediyorum; yani gördüğünüzde arkadaşlarınızı tanıma yetinizden yürüme, okuma, yazma, hesaplama, şarkı söyleme ve şaka anlatma yeteneklerinize kadar her şey.

Doğduğunuzda bu görevlerin hiçbirini yerine getiremezdiniz, bu yüzden de öğrenme dediğimiz bir süreç yoluyla tüm bu yazılımlar beyninizde programlandı. Çocukken aldığınız eğitim büyük oranda ne öğrenmeniz gerektiğine karar veren aileniz ve öğretmenleriniz tarafından tasarlanırken, giderek kendi yazılımınızı tasarlama gücünü ele geçirirsiniz. Belki de okulunuz yabancı bir dil öğrenmeye izin veriyordur: Size Fransızca ya da İspanyolca konuşturacak bir yazılım modülünü beyninize yüklemek istiyor musunuz? Tenis ya da satranç oynamayı öğrenmek istiyor musunuz? Aşçılık, avukatlık ya da eczacılık eğitimi mi almak istiyorsunuz? Yapay zekâ (YZ) ve yaşamın geleceği hakkında daha fazla şey öğrenmek için bir kitap mı okumak istiyorsunuz?

Yaşam 2.0'ın kendi yazılımını tasarlama yetisi onu Yaşam 1.0'dan çok daha zeki yapar. Yüksek zekâ hem çok fazla donanım (atomlardan yapılmış) hem de çok fazla yazılım (bitlerden yapılmış) gerektirir. İnsan donanımımızın büyük kısmının doğumdan

sonra (büyüme yoluyla) eklendiği gerçeği faydalıdır çünkü nihai boyutumuz annemizin doğum kanalının genişliğiyle sınırlı değildir. Aynı şekilde, insan yazılımımızın büyük kısmının da doğumdan sonra (öğrenme yoluyla) eklendiği gerçeği de faydalıdır çünkü nihai zekâmız da doğumda DNA'mız yoluyla bize aktarılan bilgiyle, 1.0 tarzı bir aktarımla sınırlı değildir. Doğduğum andakinden yirmi beş kat daha ağırım ve beynimdeki nöronları birbirine bağlayan sinaptik bağlantılar da doğduğum DNA'dan yüz bin kat daha fazla bilgi depolayabiliyor. Sinapslarınız tüm bilgi ve yetilerinizi, yaklaşık yüz terabaytlık bilgiyi, depolayabilirken DNA'nız en fazla bir gigabaytlık, yani yaklaşık tek bir filme yetecek kadar yere sahiptir. Yani bir çocuğun mükemmel İngilizce konuşarak ve üniversite sınavlarını geçmeye hazır biçimde doğması fiziksel olarak imkânsızdır: Böyle bir bilginin beynine önceden yüklenmiş olmasının ihtimali yoktur çünkü ebeveynlerinden aldığı temel bilgi modülü (DNA) yeterli bilgi saklama kapasitesine sahip değildir.

Kendi yazılımını tasarlama yetisi Yaşam 2.0'ı Yaşam 1.0'dan daha zeki yapmakla kalmaz, daha esnek de yapar. Eğer çevre değişirse, 1.0 yalnızca nesiller boyunca yavaşça evrimleşerek adapte olabilir. Öte yandan, Yaşam 2.0 bir yazılım güncellemesiyle neredeyse anında adapte olabilir. Mesela, sık sık antibiyotiklerle karşılaşan bakteriler nesiller boyunca evrimleşerek ilaç direnci kazanabilir ancak tek bir bakteri davranışını değiştiremez; aksine, fıstık alerjisi olduğunu öğrenen bir kız hemen davranışını değiştirerek fıstıklardan kaçınmaya başlayabilir. Bu esneklik Yaşam 2.0'a popülasyon seviyesinde ciddi bir üstünlük sağlar. İnsan DNA'mızdaki bilgi son elli bin yılda pek ciddi biçimde evrilmese de beyinlerimizde, kitaplarda ve bilgisayarlarda kolektif olarak depolanan bilgi patlama yapmıştır. Sofistike konuşma dilleriyle iletişim kurmamıza fırsat tanıyan bir yazılım modülü yükleyerek, bir insanın beyninde depolanan en faydalı bilgilerin diğer

beyinlere kopyalanabilmesini ve orijinal beyin öldükten sonra bile varlığını sürdürmesini garanti altına alıyoruz. Okuma ve yazmamızı mümkün kılan bir yazılım modülü yükleyerek, insanların ezberleyebileceğinden çok daha fazla bilgiyi depolayıp paylaşabilir hale geliyoruz. Teknoloji üretebilme kabiliyetine sahip beyin yazılımları geliştirerek (mesela bilim ve mühendislik üzerine araştırmalar yaparak), dünyadaki çoğu bilginin dünyadaki çoğu insan tarafından bir tıkla ulaşılabilir olmasını sağlıyoruz.

Bu esneklik Yaşam 2.0'ın dünyaya hükmetmesinin yolunu açmıştı. Genetik zincirlerinden kurtulan insanlığın toplam bilgisi, her bir atılım bir sonrakini mümkün kılarken artan bir ivmeyle büyümeyi sürdürdü: dil, yazı, baskı, modern bilim, bilgisayarlar, internet vs. Paylaştığımız yazılımın bu giderek hızlanan kültürel evrimi, insanlığın geleceğini şekillendiren ana güç olarak oldukça yavaş kalan biyolojik evrimimizi neredeyse alakasız kıldı.

Bugün sahip olduğumuz en güçlü teknolojilere rağmen bildiğimiz tüm yaşam formları temel olarak biyolojik donanımlarınca sınırlandırılmış olanlardır. Hiçbir şey bir milyon yıl yaşayamaz, tüm Vikipedi'yi ezberleyemez, tüm bilimi anlayamaz veya bir uzay aracı olmadan uzayda yolculuk edemez. Hiçbir şey büyük ölçüde cansız olan kozmosumuzu milyar ya da trilyon yıl boyunca serpilecek muhtelif bir biyosfere dönüştürerek evrenimizin sonunda potansiyelini gerçekleştirmesini ve tamamen uyanmasını sağlayamaz. Tüm bunlar yaşamın son bir güncelleme daha görmesini gerektiriyor, sadece yazılımını değil donanımını da dizayn edebileceği Yaşam 3.0. Diğer bir deyişle Yaşam 3.0, sonunda evrimsel zincirlerinden tamamen özgürleşmiş kendi kaderinin efendisidir.

Yaşamın üç aşaması arasındaki sınırlar hafif bulanıktır. Eğer bakteriler Yaşam 1.0 ise ve insanlar da Yaşam 2.0 ise, fareyi 1.1 olarak sınıflandırabiliriz. Pek çok şey öğrenebilirler ancak bir dil geliştirmeye ya da interneti icat etmeye yetmez bu. Dahası,

dilleri olmadığı için öğrendikleri büyük ölçüde onlarla beraber ölür ve sonraki nesillere geçmez. Benzer biçimde, bugünün insanlarının Yaşam 2.1 sayılması gerektiğini savunabiliriz. Yapay diş, diz ve kalp pili takmak gibi ufak donanım güncellemeleri yapabiliyoruz ancak on kat daha uzun olmak ya da bin kat daha büyük bir beyne sahip olmak sınırlarımızın dışında.

Özetle, yaşamın gelişimini, kendini tasarlama yetisine göre üç aşamaya ayırabiliyoruz:

- Yaşam 1.0 (biyolojik aşama): donanım ve yazılımı evrimleşir
- Yaşam 2.0 (kültürel aşama): donanımı evrimleşir, yazılımının büyük kısmını tasarlar
- Yaşam 3.0 (teknolojik aşama): donanım ve yazılımını tasarlar

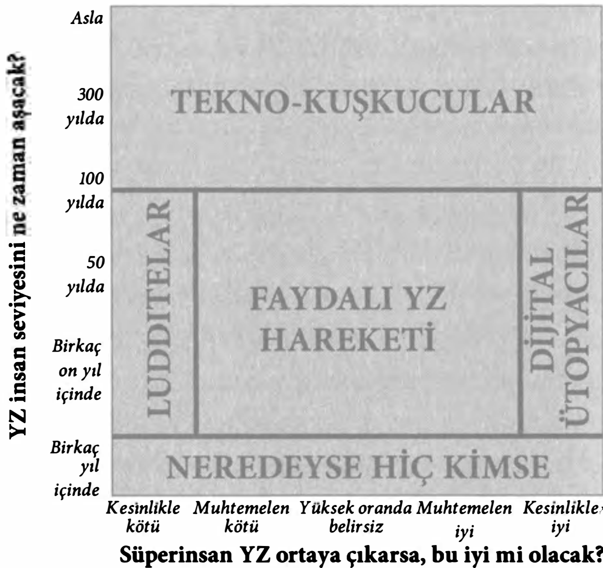
13.8 milyar yıllık kozmik evrimin ardından, dünyada gelişim önemli ölçüde hızlandı. Yaşam 1.0 yaklaşık dört milyar yıl önce geldi, Yaşam 2.0 (biz insanlar) yaklaşık yüz bin yıl önce vardık ve pek çok YZ araştırmacısı YZ'deki gelişmelerle birlikte, Yaşam 3.0'ın önümüzdeki yüzyıl içerisinde, hatta belki de bizim yaşam süremiz içinde gelebileceğini düşünüyor. Ne olacak ve bu bizim için ne anlam ifade edecek? Bu kitabın konusu bu.

İhtilaflar

Bu soru muazzam biçimde ihtilaflıdır. Dünyanın en önemli YZ araştırmacıları yalnızca tahminlerinde değil, kendinden emin iyimserlikten ciddi endişeye kadar uzanan bir skalada bulunan duygusal tepkilerinde de ciddi anlamda anlaşmazlık yaşarlar. YZ'nin ekonomik, hukuki ve askerî etkisi gibi daha kısa dönemli sorularda bile bir fikir birliği yoktur ve anlaşmazlıkları zaman aralığını genişleterek *yapay genel zekâ* (YGZ) hakkında sorular sorduğumuzda daha da artar. Özellikle de YGZ'nin insan

seviyesine ulaşıp onu geçerek Yaşam 3.0'ı ortaya çıkarması konusundaki sorularda. *Genel zekâ* satranç oynayan bir programın dar zekâsına kıyasla, öğrenmek de dâhil, aşağı yukarı her görevi gerçekleştirebilir.

İlginç biçimde, Yaşam 3.0 üzerindeki ihtilaf bir değil iki farklı soru üzerinde dönüyor: ne zaman ve ne? Ne zaman (eğer olacaksa) olacak ve bu insanlık için ne ifade edecek? Bana göre, ciddiye alınması gereken üç farklı düşünce ekolü var çünkü her biri bir dizi dünya çapında önemli uzmanı barındırıyor. Görsel 1.2'de görüldüğü gibi, onları sırasıyla ***dijital ütopyacılar***, ***tekno-kuşkucular*** ve ***faydalı-YZ hareketinin üyeleri*** olarak düşünüyorum. Size bu grupların en etkili üyelerini tanıtmama izin verin.



Görsel 1.2: Güçlü YZ (herhangi bir bilişsel görevde insanlara eş olabilecek) konusunun etrafındaki ihtilafların birçoğu iki soru üzerine odaklanır: (eğer olacaksa) ne zaman olacak ve bu insanlık için iyi bir şey mi olacak? Tekno-kuşkucular ve dijital ütopyacılar endişe etmememiz gerektiği konusunda hemfikirler ancak çok farklı sebeplerden ötürü. İlk grup insan seviyesinde

yapay genel zekânın (YGZ) öngörülebilir bir gelecekte gerçekleşmeyeceğine eminler; diğer grup ise gerçekleşeceğini düşünüyor ama iyi bir şey olacağından eminler. Faydalı YZ hareketi bu endişelerin yerinde ve faydalı olduğunu düşünüyor çünkü şu anda yapılan YZ güvenlik araştırmaları ve tartışmalar gelecekte iyi bir sonuç elde etme ihtimalini artırıyor. Ludditelar* kötü bir sonuç çıkacağına ikna olmuş durumdalar ve YZ'ye karşı geliyorlar. Bu görsel Tim Urban'ın şu makalesinden kısmen esinlenmiştir: "The AI Revolution: Our Immortality or Extinction?" *Wait But Why* (27 Ocak, 2015), http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html.

Dijital Ütopyacılar

Ben çocukken, milyarderlerin şatafatlı ve kibirli olduklarını düşünürdüm. 2008'de Google'dan Larry Page'le tanıştığımda bu basmakalıp düşüncelerim tamamen yıkıldı. Kot pantolon ve sıradan tişörtlü rahat bir giyimi olan Page, bir MIT pikniğinde kendini asla belli etmezdi. Düşünceli, yumuşak konuşma tarzı ve dostça gülümsemesi onunla konuşurken beni rahatsız etmiyor tam tersine rahatlatıyordu. 18 Haziran 2015'te Napa Vadisi'nde Elon Musk ve o zamanki eşi Talulah tarafından verilen bir davette yeniden karşılaştık ve çocuklarımızın dışkıya olan ilgileri hakkında konuştuk. Ona Andy Griffiths'in derin edebî klasiği *The Day My Butt Went Psycho* (Popomun Delirdiği Gün) isimli kitabını önerdim ve Larry hemen orada sipariş verdi. Kendime bu adamın dünya tarihine yaşamış en etkili insan olarak geçme ihtimali olduğunu hatırlatmakta zorlandım: Eğer süper zeki dijital yaşam benim yaşam sürem içinde evreni kasıp kavurursa bunun Larry'nin kararı olacağını tahmin ediyorum.

Eşlerimiz Lucy ve Meia'yla akşam yemeği yedik ve makinelerin bilinçli olup olmayacağına dair tartıştık. Ona göre bu konu dikkati asıl konudan uzaklaştıran bir tartışmaydı. O gece, kokteyllerden

* Ludditelar (Makine kırıcılar), Sanayi Devrimi esnasında makinelerin beklenen gelişmeyi vermemesi ve işlerini ellerinden alması sebebiyle makinelerini kırarak tepkilerini dile getiren bir İngiliz tekstil işçi grubu. (ç. n.)

sonra, Elon ve Larry arasında YZ'nin geleceği ve ne yapılması gerektiği üzerine uzun ve tutkulu bir tartışma başladı. Sabahın erken saatlerinde, dinleyicilerin ve tartışmaya dâhil olanların sayısı giderek artıyordu. Larry, benim *dijital ütopyacılık* olarak tanımlamak istediğim konumu tutkulu biçimde savunuyordu. Dijital yaşam, kozmik evrimde doğal ve arzu edilen bir sonraki adımdı ve dijital makineleri durdurup onları köleleştirmek yerine onları özgür bırakırsak sonucun iyi olacağından neredeyse emindi. Larry'yi dijital ütopyacılığın en etkili örneği olarak görüyorum. Yaşam eğer galaksimiz ve ötesine uzanacaksa, ki uzanması gerektiğini düşünüyordu, bunun dijital formda olması gerektiğini savunuyordu. Temel endişeleri, YZ paranoyasının dijital ütopyayı engellemesi ve/veya Google'ın "Şeytanlaşma!" sloganına ters düşecek biçimde YZ'nin ordu tarafından ele geçirilmesi idi. Elon karşı çıkıyor ve Larry'den argümanlarının detaylarını açıklamasını istiyordu. Mesela, dijital yaşamın bizim önemsediğimiz her şeyi yok etmeyeceğine dair nasıl bu kadar emin olduğunu. Bazen Larry, Elon'u "türcü" olmakla suçluyordu: karbon bazlı değil de silikon bazlı oldukları için bazı yaşam formlarını daha aşağı görmekle. Bu ilgi çekici konuları ve argümanları detaylı olarak bölüm 4'ten itibaren incelemeye başlayacağız.

Larry havuzun başındaki o ılık yaz gecesinde sayıca üstünlüğü kaybetmiş olsa da incelikli biçimde savunduğu dijital ütopyacılığın çok fazla önemli destekçisi var. Robotikçi ve fütürist Hans Moravec 1988 tarihli klasik kitabı *Mind Children* ile bir nesil boyunca dijital ütopyacıları etkilemişti. Bu gelenek mucit Ray Kurzweil tarafından devam ettirilmiş ve geliştirilmişti. Takviyeli öğrenme adıyla bilinen YZ alt dalının ileri gelen isimlerinden biri olan Richard Sutton size anlatacağım Porto Riko Konferansı'mızda dijital ütopyacılığın tutkulu bir savunmasını vermişti.

Tekno-kuşkucular

Bir diğer önemli düşünür grubu da YZ konusunda endişeli değil fakat tamamen başka bir sebepten ötürü. Süper insan YGZ inşa etmenin çok zor olduğunu ve yüzlerce yıl boyunca gerçekleşmeyeceğini, bu yüzden de şu an endişe etmenin aptalca olduğunu düşünüyorlar. Bu pozisyonu *tekno-kuşkucu* olarak adlandırıyorum; onların düşüncesini en incelikli aktaran Andrew Ng'dir: "Ölümcül robotların yükselişinden korkmak, Mars'ta nüfusun aşırı artmasından endişelenmek gibidir." Andrew, Çin'in Google'ı Baidu'nun baş mühendisidir ve Boston'da bir konferansta onunla konuştuğumda bu argümanını bana tekrar etti. Ayrıca, YZ riski hakkında endişe duymanın YZ gelişimini yavaşlatacak zararlı bir dikkat dağıtıcı olabileceğinden endişe ettiğini de söyledi. Benzer duygular Roomba robotik elektrik süpürgesi ve Baxter endüstriyel robotun arkasındaki eski MIT profesörü Rodney Brooks tarafından da dile getirilmişti. Dijital ütopyacılar ve tekno-kuşkucuların YZ hakkında endişe etmemek gerektiği konusunda hemfikir olmalarına rağmen başka neredeyse hiçbir konuda anlaşamamalarını ilginç buluyorum. Ütopyacıların birçoğu, insan seviyesi YGZ'nin önümüzdeki yirmi ila yüz yıl içinde gerçekleşeceğini düşünürken, tekno-kuşkucular bu fikri gerçeklerden bihaber olmak ve olmayacak şeyleri hayal etmek olarak görerek, kehanet edilen tekilliği "okuldaki ineklerin esrimesi" diyerek aşağılıyorlar. Rodney Brooks ile 2014 Aralık ayında bir doğum günü partisinde karşılaştığımda, bana yaşarken bunu görmeyeceğime %100 emin olduğunu söylemişti. Sonrasında, "%99 demek istemediğine emin misin?" diye yazdığım e-postayı, "Hayır %99 filan değil. %100. Kesinlikle olmayacak," diye yanıtlamıştı.

Faydalı-YZ Hareketi

Haziran 2014'te Stuart Russell'la Paris'te bir kafede tanıştığımda, ideal bir İngiliz centilmeni olarak beni etkilemişti. Güzel konuşan, düşünceli ve tatlı dilli ancak gözlerinde de maceracı bir

parıltı barındıran bu adam, bana Jules Verne'in 1873'te basılmış klasiği *80 Günde Devriâlem* romanının başkarakteri ve benim de çocukluk kahramanım olan Phileas Fogg'un modern bir versiyonu gibi gelmişti. Yaşayan en ünlü YZ araştırmacılarından biri ve konu hakkındaki standart ders kitaplarından birinin eş yazarı olmasına rağmen, alçak gönüllülüğü ve sıcakkanlılığı yüreğime su serpmişti. YZ'deki ilerlemenin bu yüzyıl içinde insan seviyesinde YGZ'yi görmemizin ciddi bir olasılık olduğuna onu nasıl ikna ettiğini anlattı ve umutlu olsa da buradan iyi bir sonucun çıkacağı kesin değildi. Başta cevaplanması gereken hayati sorular vardı ve bu sorular o kadar zordu ki onları araştırmaya şimdi başlamak zorundaydık, böylece onlara ihtiyaç duyduğumuzda cevaplara sahip olabilmeliydik.

Bugün Stuart'ın görüşleri görece ana akımdır ve dünyadaki pek çok grup onun savunduğu türde YZ-güvenlik araştırmalarını takip etmektedir. Ancak bu her zaman böyle değildi. *The Washington Post*'taki bir makale, 2015 yılını YZ-güvenlik araştırmalarının ana akım olduğu yıl olarak gösteriyordu. Bundan önce, YZ risklerini konuşmak ana akım YZ araştırmacıları tarafından sıkça yanlış anlaşılıyor ve YZ gelişimine zarar vermeyi hedefleyen Luddite korku tellallığı olarak görülüyordu. Bölüm 5'te inceleyeceğimiz üzere, Stuart'ınkilere benzer endişeler yaklaşık yarım yüzyıl önce bilgisayar öncüsü Alan Turing ve Turing'le beraber İkinci Dünya Savaşı sırasında Alman şifrelerini çözmek için çalışan matematikçi Irving J. Good tarafından dile getirilmişti. Son on yıllarda, böyle konular üzerine yapılan araştırmalar Eliezer Yudkowsky, Michael Vassar ve Nick Bostrom gibi profesyonel YZ araştırmacısı olmayan bir avuç bağımsız düşünür tarafından gerçekleştirilmişti. Başarının uzun dönemli sonuçlarını düşünmek yerine daha zeki YZ sistemleri yapma görevlerine odaklanmaya eğilimli ana akım YZ araştırmacıları üzerinde çalışmalarının pek az etkisi olmuştu.

Tanıdığım YZ araştırmacıları arasında bazı endişelere sahip olanların pek çoğu yaygaracı ve teknolojiden korkan kişiler olarak adlandırılmamak için sessiz kalmayı seçmişlerdi.

Bu kutuplaşmış durumun değişmesi gerektiğini düşünüyordum. Böylece tüm YZ camiası faydalı YZ'yi nasıl kurmak gerektiğine dair tartışmaya katılabilir ve etkileyebilirdi. Neyse ki yalnız değildim. 2014 baharında eşim Meia, fizikçi dostum Anthony Aguirre, Harvard doktora öğrencisi Viktoriya Krakovna ve Skype kurucusu Jaan Tallinn'le Future of Life Institute* (FLI; http://futureoflife.org) isimli bir kâr amacı gütmeyen organizasyon kurdum. Hedefimiz basitti: yaşamın gelecekte de sürmesini garanti altına almak ve mümkün olabildiğince harika olmasını sağlamak. Özel olarak da teknolojinin yaşama daha önce hiç olmadığı kadar gelişmesi veya kendini yok etmesi için büyük bir güç sağladığını düşünüyorduk ve ilkinin gerçekleşmesini tercih ediyorduk.

İlk toplantımız 15 Mart 2014'te bizim evde, otuz kadar öğrenci, profesör ve Boston civarından diğer düşünürlerin davet edildiği bir beyin fırtınası seansıydı. Biyoteknoloji, nükleer silahlar ve iklim değişikliğine önem vermemiz gerekse de ilk ana hedefimizin YZ güvenlik araştırmasının ana akım olmasını sağlamak olması gerektiği konusunda hemfikirdik. MIT fizikten arkadaşım, kuarkların nasıl çalıştığını ortaya çıkardığı için Nobel Ödülü kazanan Frank Wilczek, bir serbest kürsü yazısı yazarak konuya dikkat çekmeyi ve görmezden gelinmesini zorlaştırmayı teklif etti. Henüz tanışmadığım Stuart Russell'a ve fizikçi meslektaşım Stephen Hawking'e ulaşarak, onları ortak yazar olarak bana ve Frank'e katılmaya ikna ettim. Pek çok versiyon sonunda, yazımız *The New York Times* ve pek çok ABD gazetesinden ret cevabı aldı

* (İng.) Yaşamın Geleceği Enstitüsü. (ç. n.)

biz de onu *Huffington Post* blog hesabımdan yayımladık. Muhteşem bir sürpriz olarak Arianna Huffington bana e-posta attı ve "Bu yazıyı aldığımıza çok sevindik! Bir numaradan yayımlayacağız!" dedi ve ana sayfanın en başına koyulan bu yazı, yılın geri kalanında YZ güvenliği üzerine bir medya görünürlüğü dalgası yarattı. Elon Musk, Bill Gates ve diğer teknoloji liderleri de buna katkı sundu. Nick Bostrom'un kitabı *Superintelligence* o sonbahar raflardaki yerini aldı ve artan tartışmayı ateşledi.

FLI faydalı YZ kampanyamızın bir sonraki hedefi de dünyanın ileri gelen YZ araştırmacılarını yanlış anlaşılmaların düzeltilebileceği, fikir birliğinin kurulabileceği ve yapıcı planların yapılabileceği bir konferansta toplamaktı. Böylesine ünlü bir kalabalığı özellikle de tartışmalı bir konu olunca, tanımadıkları insanların düzenlediği bir konferansa gelmeye ikna etmenin zor olacağını biliyorduk. Biz de elimizden geleni ardımıza koymadık: Medyanın katılmasını engelledik, ocak ayında (Porto Riko'da) bir tatil köyünü ayarladık, katılımı ücretsiz yaptık (Jaan Tallinn'in cömertliği sayesinde) ve olabilecek en az yaygaracı ismi koyduk: "YZ'nin Geleceği: Fırsatlar ve Zorluklar." Daha da önemlisi, sayesinde organizasyon komitesini hem akademiden hem de endüstriden bir grup YZ liderini –YZ'nin insanları Go oyununda bile yenebileceğini gösteren Google'ın DeepMind'ından Demis Hassabis de dâhil olmak üzere– kapsayacak şekilde büyüttüğümüz Stuart Russell'la güçlerimizi birleştirdik. Demis'i tanıdıkça, YZ'yi yalnızca güçlü yapmak değil, faydalı yapmak da istediğini anlamaya başladım.



Görsel 1.3: Ocak 2015 Porto Riko Konferansı YZ ve alakalı alanlardan olağanüstü bir grup araştırmacıyı bir araya getirdi. Arka sıra, soldan sağa: Tom Mitchell, Seán Ó hÉigeartaigh, Huw Price, Shamil Chandaria, Jaan Tallinn, Stuart Russell, Bill Hibbard, Blaise Agüera y Arcas, Anders Sandberg, Daniel Dewey, Stuart Armstrong, Luke Muehlhauser, Tom Dietterich, Michael Osborne, James Manyika, Ajay Agrawal, Richard Mallah, Nancy Chang, Matthew Putman. Diğer ayaktakiler, soldan sağa: Marilyn Thompson, Rich Sutton, Alex Wissner-Gross, Sam Teller, Toby Ord, Joscha Bach, Katja Grace, Adrian Weller, Heather Roff- Perkins, Dileep George, Shane Legg, Demis Hassabis, Wendell Wallach, Charina Choi, Ilya Sutskever, Kent Walker, Cecilia Tilli, Nick Bostrom, Erik Brynjolfsson, Steve Crossan, Mustafa Suleyman, Scott Phoenix, Neil Jacobstein, Murray Shanahan, Robin Hanson, Francesca Rossi, Nate Soares, Elon Musk, Andrew McAfee, Bart Selman, Michele Reilly, Aaron VanDevender, Max Tegmark, Margaret Boden, Joshua Greene, Paul Christiano, Eliezer Yudkowsky, David Parkes, Laurent Orseau, JB Straubel, James Moor, Sean Legassick, Mason Hartman, Howie Lempel, David Vladeck, Jacob Steinhardt, Michael Vassar, Ryan Calo, Susan Young, Owain Evans, Riva-Melissa Tez, János Kràmar, Geoff Anders, Vernor Vinge, Anthony Aguirre. Oturanlar: Sam Harris, Tomaso Poggio, Marin Soljačić, Viktoriya Krakovna, Meia Chita-Tegmark. Kamera arkasında: Anthony Aguirre (yanında oturan insan seviyesinde zekâ tarafında photoshop ile eklendi).



Görsel 1.4: Medya Elon Musk'ı sık sık YZ camiasıyla arası bozuk şekilde lanse etse de YZ güvenlik araştırmasının gerekli olduğuna dair geniş bir fikir birliği mevcut. Burada, 4 Ocak 2015'te, Yapay Zekânın İlerlemesi Derneği başkanı Tom Dietterich, kısa süre önce Elon'un fonlamaya söz verdiği yeni YZ güvenlik araştırma programı hakkında Elon'un heyecanını paylaşırken görünüyor. FLI kurucuları Meia Chita-Tegmark ve Viktoriya Krakovna arkalarında saklanıyor.

Sonuç zihinlerin olağanüstü bir toplantısıydı (Görsel 1.3). Üst düzey iktisatçılar, hukukçular, teknoloji liderleri (Elon Musk da dâhil olmak üzere) ve diğer düşünürler (Bölüm 4'ün ana odağı olacak "tekillik" terimini ortaya atan Vernor Vinge de dâhil olmak üzere) YZ araştırmacılarına katılmışlardı. Sonuç en iyimser beklentilerimizi bile aştı. Belki güneş ışığı ve şarabın bir araya gelmesindendi ya da belki de doğru zamanı bulmuştuk: İhtilaflı konuya rağmen, olağanüstü bir fikir birliği ortaya çıktı. Bunun sonucunu YZ'deki önemli herkesin dâhil olduğu sekiz bin kişi tarafından imzalanmış bir açık mektupta birleştirdik.[1] Mektubun esası YZ'nin hedefinin yeniden tanımlanması gerektiğiydi. Amaç güdümsüz bir zekâ değil faydalı bir zekâ yaratmak olmalıydı. Mektup ayrıca konferans katılımcılarının anlaştığı detaylı bir araştırma konuları listesinin bu amacı ileri taşıyacağından da

bahsediyordu. Faydalı YZ hareketi ana akım olmaya başlamıştı. Takip eden ilerlemeyi bu kitabın ilerleyen kısımlarında göreceğiz.

Konferanstan çıkan diğer bir önemli ders şuydu: YZ'nin başarısından ortaya çıkan sorular yalnızca entelektüel olarak baş döndürücü değildi, ayrıca ahlaki olarak hayatiydi de çünkü seçimlerimiz yaşamın geleceğini tümden değiştirme potansiyeline sahipti. İnsanlığın geçmiş seçimlerinin ahlaki önemi bazen çok büyüktü ama her zaman sınırlıydı: En büyük salgınlardan kurtulduk ve en büyük imparatorluklar bile eninde sonunda parçalandı. Geçmiş nesiller bundan güneşin ertesi gün doğacağı kadar emindiler, bu yüzden de geleceğin insanlarının da yapacağı gibi, fakirlik, hastalık ve savaş gibi tekrar eden musibetlerle mücadele ediyorlardı. Fakat Porto Riko'daki konuşmacıların bir kısmı bu sefer her şeyin biraz farklı olabileceğini ileri sürdüler. Onlara göre ilk kez, bu felaketleri –ya da insanlığın kendini- bitirebilecek kadar güçlü bir teknoloji inşa edebilirdik. Dünyada ya da ötesinde, daha önce hiç olmadığı kadar gelişen toplumlar ya da asla devrilemeyecek kadar güçlü, Kafkaesk bir küresel gözetim devleti oluşturabilirdik.

Yanlış Kanılar

Porto Riko'dan ayrıldığımda, YZ'nin geleceği hakkında orada yaptığımız tartışmanın devam etmesi gerektiğine ikna olmuştum çünkü bu, çağımızın en önemli tartışması idi.* Hepimizin kolektif geleceği hakkında bir tartışmaydı bu, o yüzden de YZ

* YZ tartışması hem etki hem de aciliyet bakımında önemlidir. Elli ila iki yüz yıl arasında ortalığı kasıp kavurabilecek iklim değişikliğine kıyasla, pek çok uzman YZ'nin on yıllar içinde daha büyük etki yaratmasını bekliyor. Hem de bize iklim değişikliğinin etkilerini hafifletebilecek bir teknoloji verme potansiyeli var. Savaşlar, terörizm, işsizlik, açlık, göç ve sosyal adalet konularına kıyasla, YZ'nin yükselişinin toplam etkisi daha çok olacak. Gerçekten de tüm bu konuları nasıl domine edebileceğini, iyi ya da kötü yönde, bu kitapta göreceğiz.

araştırmacılarıyla sınırlı kalamazdı. Bu kitabı yazma sebebim de bu: Bu kitabı, siz sevgili okuyucuların da bu tartışmaya katılması ümidiyle yazdım. Ne tür bir gelecek istiyorsunuz? Ölümcül otonom silahlar yaratmalı mıyız? İş otomasyonu konusunda neler olmasını isterdiniz? Bugünün çocuklarına nasıl kariyer tavsiyeleri verirsiniz? Yeni işlerin eskilerinin yerini almasını mı tercih edersiniz yoksa herkesin bolca boş vaktinin olduğu ve makine üretimi servetin tadını çıkardığı işsiz bir dünyayı mı? Yolun daha da ilerisinde, Yaşam 3.0'ı yaratıp kozmosun ötesine yaydığımızı görmek ister miydiniz? Zeki makineler bizim yerimizi mi alacak, bizle beraber mi var olacak yoksa bizimle birleşecek mi? Yapay zekâ çağında insan olmak ne anlama gelecek? Ne anlama gelmesini istersiniz ve geleceğin bu yönde olması için ne yapabiliriz?

Bu kitabın amacı tartışmaya katılmanıza yardımcı olmak. Dediğim gibi dünyanın ileri gelen uzmanlarının uzlaşamadığı baş döndürücü ihtilaflar var. Ancak insanların birbirini yanlış anladığı ve ayrı telden çaldıkları birçok sıkıcı sözde ihtilaf örneğine de tanıklık ettim. Yanlış anlaşılmalara değil de ilgi çekici ihtilaflar ve açık sorulara odaklanmamıza yardımcı olması için, bazı en sık görülen yanlış kanıları ortadan kaldırmakla başlayalım.

"Yaşam," "zekâ" ve "bilinç" gibi terimlerin gündelik kullanımı için birçok rakip tanım bulunur ve çoğu yanlış kanı da insanların bir kelimeyi iki farklı şekilde kullandıklarını fark etmemelerinden ileri gelir. Bu tuzağa düşmememiz için, bu kitaptaki ana terimleri nasıl kullandığımı gösteren bir tabloyu sizlere sunuyorum. Bu tanımların bir kısmı ancak ilerleyen bölümlerde etraflıca anlatılacak ve açıklanacak. Lütfen tanımlarımın başkalarınınkinden daha iyi olduğunu iddia etmediğimi göz önünde bulundurun. Yalnızca söylediğim şeylerde açık olarak karmaşaya mahal vermek istemiyorum. Genel olarak insan merkezci önyargıya düşmeyen ve makinelere olduğu kadar insanlara da uygulanabilen genel

tanımları seçtiğimi göreceksiniz. Lütfen tabloyu şimdi okuyun ve bu kelimeleri nasıl kullandığıma dair kafanız karışırsa geri gelip kontrol edin. Özellikle de Bölüm 4-8 arasında.

Terminoloji Tablosu	
Yaşam	Kendi karmaşıklığını sürdürebilen ve kendini kopyalayabilen süreç
Yaşam 1.0	Kendi donanım ve yazılımı evrim geçiren yaşam (biyolojik aşama)
Yaşam 2.0	Kendi donanımı evrim geçiren ama yazılımının büyük kısmını tasarlayan yaşam (kültürel aşama)
Yaşam 3.0	Kendi donanım ve yazılımını tasarlayan yaşam (teknolojik aşama)
Zekâ	Karmaşık görevleri yerine getirme yetisi
Yapay Zekâ (YZ)	Biyolojik olmayan zekâ
Dar zekâ	Satranç oynamak ya da araba sürmek gibi dar bir görev setini yerine getirebilme yetisi
Genel zekâ	Öğrenmek de dâhil olmak üzere tüm hedefleri yerine getirme yetisi
Evrensel zekâ	Veri ve kaynaklara erişimi olduğunda genel zekâ kazanma yetisi
[İnsan seviyesi] Yapay Genel Zekâ (YGZ)	En azından insanlar kadar herhangi bir bilişsel görevi yerine getirme yetisi
İnsan seviyesi YZ	YGZ
Güçlü YZ	YGZ
Süper zekâ	İnsan seviyesinin çok ötesinde genel zekâ
Uygarlık	Zeki yaşam formlarının etkileşim içinde olduğu grup
Bilinç	Öznel deneyim
Qualia	Öznel deneyimin tekil anları

Etik	Nasıl davranmamız gerektiğini belirleyen ilkeler
Teleoloji/Erek-bilim	Şeyleri sebeplerinden ziyade hedefleri veya amaçlarıyla açıklamak
Hedefe yönelik davranış	Sebebiyle değil etkisiyle daha kolay açıklanan davranış
Hedef sahibi olmak	Hedefe yönelik davranış gösterme
Amaca sahip olmak	Kendinin ya da başka bir varlığın hedeflerine hizmet etmek
Dost YZ	Hedefleri bizimkilerle aynı eksende olan süper zekâ
Cyborg	İnsan-makine melezi
Zekâ patlaması	Süper zekâya hızla yol veren tekrar eden öz gelişim
Tekillik	Zekâ patlaması
Evren	Büyük Patlama'dan bu yana geçen 13.8 milyar yıl boyunca ışığın bize ulaşacak zaman bulabildiği uzay kesimi

Tablo 1.1: YZ hakkındaki pek çok yanlış kanı insanların yukarıdaki kelimeleri başka anlamlarda kullanmasından kaynaklanır. Bu kitapta hangi anlamla kullandığımı buradan görebilirsiniz (Bu tanımların bir kısmı ancak ilerleyen bölümlerde etraflıca anlatılacak ve açıklanacak).

Terminoloji üzerine kafa karışıklığına ek olarak, pek çok YZ tartışmasının basit yanlış kanılar yüzünden yoldan çıktığını gördüm. Şimdi en sık rastlananlarını netleştirelim.

Zaman Çizelgesi Efsaneleri

İlki Görsel 1.2'deki zaman çizelgesiyle ilgili: Makinelerin büyük ölçüde insan seviyesi YGZ'nin yerini alması için ne kadar zaman var? Bunun net bir şekilde cevabını bildiğimizin sanılması sık rastlanan yanlış kanılardan biridir.

Bir diğer popüler efsane de süper insan YGZ'yi bu yüzyıl elde edeceğimizin kesin olduğuna inanılmasıdır. Aslına bakarsanız, tarih bunun gibi teknolojik abartılarla doludur. Şimdiye kadar sahip olacağımıza dair sözler verilen o füzyon santralleri ve uçan arabalar nerede? YZ de geçmişte defalarca abartıldı, hem de bazen alanın kurucuları tarafından, mesela "yapay zekâ" terimini bulan John McCarthy, Marvin Minsky, Nathaniel Rochester ve Claude Shannon, Taş Devri bilgisayarlarıyla iki ay içerisinde neler gerçekleştirilebileceğine dair şu abartılı iyimser tahmini yazmışlardı: "Dartmouth College'da, 1956 yazı boyunca iki aylığına on kişilik bir yapay zekâ çalışması yapmayı planlıyoruz... Bu çalışmada makinelerin dili kullanmasını, soyutlamalar ve kavramlar kurmasını, yalnızca insanların yapabileceği problemleri çözmesini ve kendilerini geliştirmesini nasıl sağlayabileceğimizi bulmaya çalışacağız. Özenle seçilmiş bir grup bilim insanı bir yaz boyunca beraber çalışırsa bunlardan biri ya da daha fazlası konusunda önemli bir atılım yapabileceğimizi düşünüyoruz."

Öte yandan, bir başka efsane de süper insan YGZ'yi bu yüzyılda elde *edemeyeceğimize* dairdir. Araştırmacılar süper insan YGZ'den ne kadar uzakta olduğumuza dair geniş çaplı tahminler yapmışlardır ancak böylesine tekno-kuşkucu tahminlerin iç karartıcı geçmiş performanslarını düşündüğümüzde, emin biçimde bu olasılığın bu yüzyılda sıfır olduğunu da söyleyemeyiz. Mesela, muhtemelen zamanının en büyük nükleer fizikçilerinden biri olan Ernest Rutherford 1933 yılında –Leo Szilard'ın nükleer zincirleme tepkimesini buluşundan yirmi dört saatten kısa süre önce– nükleer enerjinin "boş laf" olduğunu söylemişti. 1956 yılında Kraliyet Astronomu Richard Woolley uzay yolculuğu hakkında tartışmalar hakkında "dev abartı" demişti. Bu efsanenin en uç hali, süper insan YGZ'nin fiziksel olarak imkânsız olduğu için asla gelmeyeceğini söylemektir. Ancak fizikçiler beynin güçlü

bir bilgisayar gibi hareket edecek şekilde düzenlenmiş kuarklar ve elektronlardan oluştuğunu ve daha zeki kuark bütünleri inşa etmekten bizi alıkoyan hiçbir fizik kanunu olmadığını biliyorlar.

YZ araştırmacılarına en az %50 ihtimalle ne zaman insan seviyesi YGZ'ye sahip olacağımızı soran bir dizi anket oldu ve tüm bu anketler aynı sonucu işaret ediyordu: Dünyanın ileri gelen uzmanları anlaşamıyordu yani bilmiyoruz. Mesela, Porto Riko YZ Konferansı'ndaki YZ araştırmacılarına yapılan böyle bir ankette, ortalama (medyan) cevap 2055 yılıydı ancak bazı araştırmacılar yüzlerce ya da daha fazla yıl sonrası gibi cevaplar verdiler.

Ayrıca bununla ilgili bir başka efsane daha var. YZ hakkında endişe duyan insanların bunun yalnızca birkaç yıl sonra gerçekleşeceğini düşünüyor olmalarına dair bir efsane bu. Aslında süper insan YGZ konusunda endişelenen kişilerin pek çoğu bunun en azından on yıllarca ötede olduğunu düşünüyor. Ancak ileri sürdükleri düşünce, bunun bu yüzyılda olmayacağından %100 *emin* olmadığımız için, ihtimallere hazırlık yapmak amacıyla güvenlik araştırmalarına şimdiden başlamanın zekice olduğu. Bu kitapta göreceğimiz üzere, pek çok güvenlik problemi o kadar zor ki çözmesi on yıllar sürebilir, bu yüzden de Red Bull içen bazı programcılar insan seviyesi YGZ'yi çalıştırmaya başlamadan bir gece önce bunları düşünmektense onları araştırmaya şimdi başlamak çok daha ihtiyatlı olur.

Efsane 2100 yılında süper zekâ kaçınılmaz olacak Efsane 2100 yılında süper zekânın gelmesi imkânsız	Gerçek On yıllar içinde de yüzyıllar içinde de olabilir ya da hiç olmayabilir. Tüm uzmanlar farklı düşünüyor yani bilmiyoruz.
Efsane Yalnızca Ludditelar YZ hakkında endişeleniyor	Gerçek En önemli YZ araştırmacılarının birçoğu endişeli
Efsane endişe YZ şeytanlaşabilir Efsane endişe YZ bilinçlenebilir	Gerçek endişe YZ yetkinleşip bizim hedeflerimizden farklı eksende hedeflere sahip olabilir YZ Biz
Efsane Robotlar endişelerin merkezinde	Gerçek Farklı eksende zekâ endişelerin merkezinde: kimseye ihtiyaç duymaz, yalnızca bir internet bağlantısı gerekir
Efsane YZ insanları kontrol edemez	Gerçek Zekâ kontrole olanak sağlar: Biz daha zeki olduğumuz için kaplanları kontrol ediyoruz
Efsane Makinelerin hedefleri olamaz	Gerçek Isı güdümlü füzelerin bir hedefi vardır
Efsane endişe Süper zekâ yalnızca birkaç yıl ötede PANİK!	Gerçek endişe En az on yıllarca ötemizde ancak bunu güvenli hale getirmek de o kadar zaman alabilir ÖNCEDEN PLANLA!

Görsel 1.5: Süper zeki YZ hakkında sık rastlanan efsaneler.

İhtilaf Efsaneleri

Bir diğer sık rastlanan yanlış kanı da YZ hakkında endişeleri olan ve YZ güvenlik araştırmalarını destekleyen tek insanların YZ hakkında pek bir şey bilmeyen Ludditelar olduğudur. Stuart

Russell bunu Porto Riko'daki konuşmasında dile getirdiğinde dinleyiciler sesli biçimde güldüler. Alakalı başka bir yanlış kanı da YZ güvenliği araştırmalarını desteklemenin cidden ihtilaflı olduğudur. Aslında YZ güvenliği araştırması için alçak gönüllü bir yatırı,mı desteklemek için insanların risklerin yüksek olduğuna ikna olmalarına gerek yok, ihmal edilemez olduğuna inanmaları yeterli. Tıpkı ev sigortasına yapılan küçük bir yatırımın evde yangın çıkmasının ihmal edilemez olasılığıyla gerekçelendirilmesi gibi.

Benim şahsi analizim medyanın YZ güvenliği tartışmasını olduğundan daha ihtilaflı gösterdiği yönündedir. Sonuçta korku gazetenin satmasına neden olur ve yaklaşan kıyameti ilan etmek için bağlam dışı alıntılar kullanan makaleler detaylı ve dengeli olanlardan daha çok tık alır. Sonuç olarak da karşısındakinin görüşünü yalnızca medya alıntılarından bilen iki insan, normalde olduğundan daha çok ayrıştıklarını düşünme eğilimindedir. Mesela, Bill Gates'in görüşüne dair tek bilgisi İngiliz bulvar gazetelerinin birinden gelen bir tekno-kuşkucu, hatalı bir biçimde Gates'in süper zekânın çok yakında geleceğine inandığını düşünebilir. Benzer bir şekilde, faydalı-YZ hareketinde olup Andrew Ng'ın pozisyonu hakkında yalnızca yukarıda verdiğimiz Mars'ın aşırı kalabalıklaşması alıntısını bilen bir kimse, YZ güvenliğini hiç önemsemediğini düşünebilir. Fakat ben şahsen biliyorum ki önemsiyor. İşin aslı, zaman çizelgesi tahminleri daha uzun olduğu için, doğal olarak kısa dönem YZ sorunlarını uzun dönem sorunların üzerinde tutma eğilimi gösteriyor.

Riskler Hakkında Efsaneler

Daily Mail'de şu başlığı görünce gözlerimi devirdim:[2] "Stephen Hawking Robotların Yükselişinin İnsanlık için Felaket Olabileceği Konusunda Uyarılarda Bulundu." Buna benzer daha kaç tane

makale gördüğümü hatırlamıyorum. Genelde, bu makalelerin yanında silah taşıyan şeytani görünümlü bir robot resmi olur ve bilinçli ve/veya şeytani oldukları için robotların ayaklanıp bizi öldürmesinden endişe etmemiz gerektiği salık verilir. Öte yandan bu makaleler aslında oldukça etkileyicidir çünkü YZ meslektaşlarımın endişe *etmediği* senaryoyu kısa ve öz biçimde özetler. Bu senaryo üç ayrı yanlış kanıyı içinde barındırır: sırasıyla, *bilinç, şeytanilik ve robot* endişeleri.

Arabayla bir yolda gidiyorsunuz diyelim, renkler, sesler ve diğer şeylerden oluşan öznel bir deneyim yaşarsınız. Peki, sürücüsüz arabalar da bu deneyimi yaşar mı? Sürücüsüz bir araba olmak bir şeyler hissettirir mi yoksa bu hiçbir öznel deneyimi olmayan bilinçsiz bir zombi olmak gibi midir? Bu bilinç gizemi kendi başına ilginç olsa da, zaten Bölüm 8'i ona ayırıyoruz, YZ riskiyle alakalı değildir. Eğer sürücüsüz bir araba size çarparsa, onun öznel olarak bilinçli hissedip hissetmemesinin bir önemi yoktur. Aynı şekilde, biz insanları etkileyecek olan şey de öznel olarak nasıl hissettiği değil, süper zeki YZ'nin ne *yaptığı* olacak.

Makinelerin şeytanileşmesi korkusu da başka bir dikkat dağıtıcı unsur. Gerçek endişe kötü niyet değil, yeterliktir. Süper zeki bir YZ, tanımı gereği hedefleri ne olursa olsun onları gerçekleştirmekte çok iyidir. Bu yüzden de hedeflerinin bizimkilerle aynı doğrultuda olduğundan emin olmalıyız. Sırf kötülük olsun diye karıncaların üzerine basan bir karınca düşmanı değilsinizdir ama bir hidroelektrik yeşil enerji projesinin başındaysanız ve sular altında kalacak bölgede karınca yuvası varsa, karıncalara yazık olur. Faydalı-YZ hareketi, insanlığı o karıncaların yerine koymaktan kaçınmaya çalışıyor.

Bilinç yanılgısı, makinelerin hedefleri olamayacağı mitine dayanıyor. Makinelerin hedefe yönelik davranış gösterme anlamında elbette hedefleri olabilir: Isı güdümlü bir füzenin davranışı

en basitçe hedefi vurmak olarak açıklanabilir. Eğer hedefi sizinkilerle aynı doğrultuda olmayan bir makine sizi tehdit ederse, tam olarak bu dar anlamda sizi rahatsız eden şey onun hedefleri olur, bilinçli olup olmadığı ya da bir amaç hissi deneyimleyip deneyimlemediği değil. Eğer ısı güdümlü füzeler sizin peşinizde olsaydı, "Endişe duymuyorum çünkü makinelerin hedefleri olamaz!" demezdiniz muhtemelen.

Korku tellallığı yapan gazeteler tarafından şeytanlaştırılan Rodney Brooks ve diğer robotik öncülerinin duygularını paylaşıyorum çünkü bazı gazeteciler takıntılı bir biçimde robotlara takıyor ve makalelerini parlak kırmızı gözleriyle şeytani görünen metal canavarlarla dolduruyorlar. Aslına bakarsanız, faydalı-YZ hareketinin ana endişesi robotlarla değil zekânın kendisiyledir. Özel olarak da hedefleri bizimle aynı doğrultuda olmayan zekâyla. Bize sorun yaratmak için yanlış doğrultudaki bir zekânın robotik bir vücuda ihtiyacı yoktur, yalnızca internet bağlantısı yeterlidir. Bölüm 4'te böyle bir şeyin finans marketlerini nasıl geride bırakabileceğini, insan araştırmacılardan nasıl daha iyi icatlar yapabileceğini, insan liderleri nasıl manipüle edebileceğini ve anlamakta bile güçlük çekeceğimiz silahlar geliştirebileceğini inceleyeceğiz. Robot yapmak fiziksel olarak imkânsız olsaydı bile, süper zeki ve süper zengin bir YZ kolayca bir sürü insana farkında olmadan istediğini yaptırabilirdi, tıpkı William Gibson'ın bilimkurgu romanı *Neuromancer*'da olduğu gibi.

Robot yanılgısı da makinelerin insanları kontrol edemeyeceği mitine dayanır. Zekâ, kontrolü mümkün kılar: İnsanlar kaplanları daha güçlü oldukları için değil daha zeki oldukları için kontrol edebilirler. Bu da gezegenin en zekisi konumunu birilerine bırakırsak, kontrolü de bırakabileceğimiz anlamına gelebilir.

Görsel 1.5 bu yaygın yanlış kanıları özetliyor. Böylece bunları aradan çıkarabilir ve pek çok geçerli ihtilaflar üzerinde dostları-

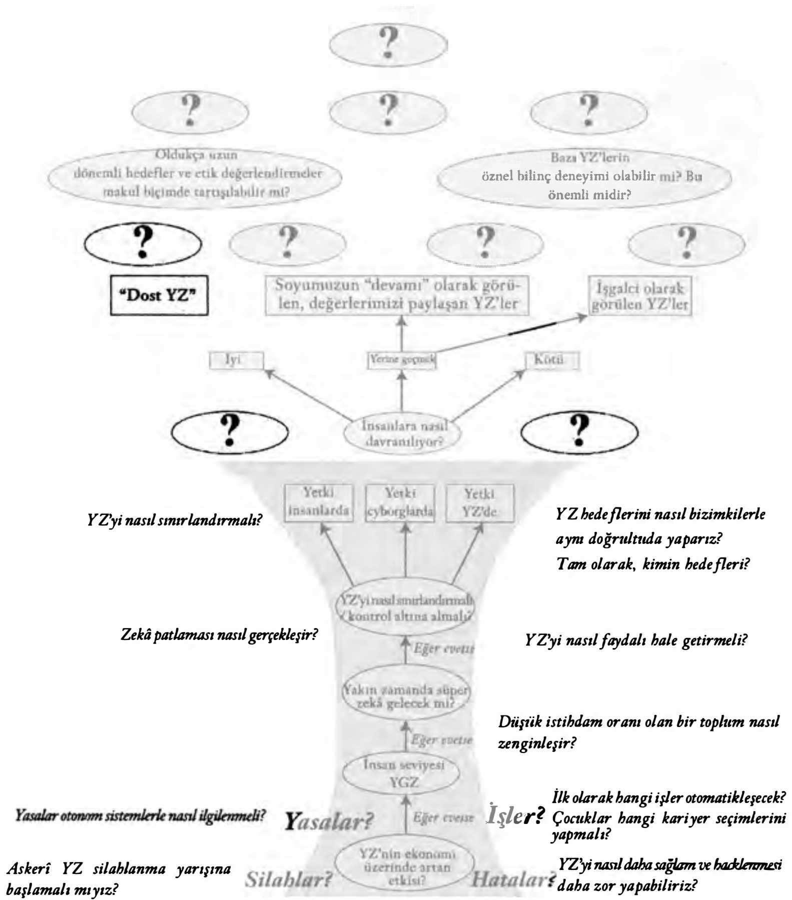
mız ve meslektaşlarımızla yaptığımız tartışmalara odaklanabiliriz. Göreceğimiz üzere de bu ihtilaflardan da oldukça çok var!

Önümüzdeki Yol

Kitabın geri kalan kısmında, sizinle beraber YZ ile yaşamın geleceğini keşfedeceğiz. Bu zengin ve çok yönlü konuya, önce kavramsal ve kronolojik olarak yaşamın tüm hikâyesini keşfederek, sonra da istediğimiz geleceği yaratmak için ne gibi adımların atılması gerektiğini, hedefleri ve anlamı inceleyerek, organize bir şekilde girelim.

Bölüm 2'de zekânın temellerini ve görünüşte zekâya sahip olmayan bir maddenin hatırlayacak, işlem yapacak ve öğrenecek şekilde nasıl yeniden düzenlenebileceğini göreceğiz. Geleceğe doğru ilerledikçe, hikâyemiz belirli ana sorulara verilen cevaplarca şekillenen pek çok senaryoyla dallanacak. Görsel 1.6 zaman geçtikçe ve potansiyel olarak çok daha gelişmiş bir YZ'ye doğru ilerledikçe karşılaşacağımız temel soruları özetliyor.

Şu anda bir YZ silah yarışına başlayıp başlamama seçeneğiyle ve geleceğin YZ sistemlerini sağlam ve hatasız nasıl yapacağımıza dair sorularla yüz yüzeyiz. Eğer YZ'nin ekonomik etkisi artmayı sürdürürse, yakında otomatikleşecek işlerden kaçınmaları için çocuklara vereceğimiz kariyer tavsiyelerinin ne olacağına ve yasalarımızı nasıl modernize edeceğimize karar vermemiz gerekecek. Bunun gibi kısa dönemli soruları Bölüm 3'te cevaplamaya çalışacağız.



Görsel 1.6: Hangi YZ soruları YZ'nin ne kadar ilerlediğine ve geleceğimizin hangi yoldan gideceğine göre ilgi alanımızda olacak?

Eğer YZ gelişimi insan seviyelerine ilerlerse, kendimize bunun faydalı olmasını nasıl sağlayacağımızı ve iş olmadan zenginleşen bir boş vakit toplumu yaratabilir miyiz veya yaratmalı mıyız diye de sormamız gerekir. Bu da YGZ'yi insan seviyelerinin ötesine zekâ patlamasının mı yoksa yavaş ama istikrarlı bir büyümenin mi taşıyacağı sorusunu akla getirir. Buna benzer, distopikten ütopiğe

senaryolardan birçoğunu Bölüm 4'te, sonucunda olabilecekleri de Bölüm 5'te inceleyeceğiz. Yetki kimde? İnsanlar, YZ ya da cyborglarda mı? İnsanlara iyi mi kötü mü davranılıyor? Bizim yerimizi birileri mi alıyor ve alıyorsa, yerimizi alanları işgalciler olarak mı yoksa soyumuzun devamı olarak mı görüyoruz? Bölüm 5 senaryolarından hangilerini tercih edeceğiniz konusunda meraktayım! Bir web sitesi açtım, http://AgeOfAi.org, burada görüşlerinizi paylaşabilir ve tartışmaya katılabilirsiniz.

Son olarak, Bölüm 6'da ironik biçimde önceki bölümlerden daha net sonuçlar çıkarabildiğimiz bir zaman dilimine, gelecekte milyarlarca yıl ileriye gidiyoruz çünkü kozmosumuzdaki yaşamın sınırları zekâ tarafından değil, fizik yasaları tarafından belirleniyor.

Zekâ tarihini keşfettikten sonra, kitabın kalan kısmını hangi geleceği hedefleyeceğimizi ve oraya nasıl varacağımızı irdelemeye ayıracağız. Gerçekleri anlam ve amaç sorularına bağlayabilmek için, Bölüm 7'de hedeflerin, Bölüm 8'de de bilincin fiziksel temelini inceleyeceğiz. En son olarak da Sonsöz bölümünde, istediğimiz geleceği yaratmak için şu an neler yapılabileceğini tartışacağız.

Eğer atlaya atlaya okumayı seven bir okuyucuysanız, ilk bölümdeki ve bir sonraki bölümün başındaki terminoloji ve tanımları sindirdiğiniz takdirde çoğu bölüm başlı başına okunabilir. Eğer bir YZ araştırmacısıysanız, isterseniz girişindeki zekâ tanımları dışında tüm Bölüm 2'yi atlayabilirsiniz. Eğer YZ'de yeniyseniz, Bölüm 2 ve 3 size Bölüm 4-6'nın neden imkânsız bilimkurgu olarak geçiştirilemeyeceğine dair argümanlar sunacak. Görsel 1.7 çeşitli bölümlerin gerçeklere dayalı olmaktan kuramsal olmaya uzanan spektrumda nereye denk düştüğünü gösteriyor.

Baş döndürücü bir yolculuk bizi bekliyor. Haydi başlayalım!

		Kısa Bölüm Adı	Konu	Durum
		Giriş: Omega Ekibinin Hikâyesi	Düşündürücü bir hikâye	Oldukça kurgusal
Zekânın tarihi	1	Tartışma	Ana fikirler, terminoloji	Pek kurgusal değil
	2	Madde Zekâ Kazanıyor	Zekânın temelleri	
	3	YZ, Ekonomi, Silahlar ve Hukuk	Yakın gelecek	
	4	Zekâ Patlaması?	Süper zekâ senaryoları	Oldukça kurgusal
	5	Sonuç	Sonraki 10000 yıl	
	6	Kozmik Birikimimiz	Sonraki milyarlarca yıl	
Anlamın tarihi	7	Hedefler	Hedefe yönelik davranışın tarihi	Pek kurgusal değil
	8	Bilinç	Doğal & yapay bilinç	Kurgusal
		Sonsöz: FLI Takımının Hikâyesi	Ne yapmalıyız?	Pek kurgusal değil

Görsel 1.7: Kitabın yapısı

SONUÇ:

- Karmaşıklığını ve kendini kopyalama özelliğini sürdürebilen bir süreç olarak tanımlanan yaşam üç aşamada gelişebilir: yazılım ve donanımının evrimleştiği biyolojik aşama (1.0), yazılımını (öğrenme yoluyla) tasarlayabilen kültürel aşama (2.0) ve donanımını da tasarlayabilen, kaderinin efendisi olan teknolojik aşama (3.0).
- Yapay zekâ, Yaşam 3.0'ın bu yüzyılda gelmesini sağlayabilir ve hangi geleceği hedeflememiz gerektiği ile buna nasıl ulaşılabileceği üzerine baş döndürücü bir tartışma başlamıştır. Bu tartışmada üç ana grup mevcut: tekno-kuşkucular, dijital ütopyacılar ve faydalı-YZ hareketi.
- Tekno-kuşkucular süper insan YGZ inşa etmenin yüzlerce yıl boyunca gerçekleşmeyecek kadar zor olduğunu düşündükleri için bu konuda (ve Yaşam 3.0 konusunda) endişe etmenin anlamsız olduğunu düşünüyorlar.
- Dijital ütopyacılar bu yüzyılda gerçekleşebileceğini düşünüyor ve tüm içtenlikleriyle Yaşam 3.0'ı kozmik evrimin doğal ve arzu edilen bir sonraki adımı olarak karşılıyorlar.
- Faydalı-YZ hareketi de bu yüzyılda olma ihtimalini yüksek görüyor ancak iyi bir sonucun garanti olmadığını, ciddi YZ-

güvenlik araştırmalarıyla garanti altına alınması gerektiğini söylüyorlar.

- Dünyanın ileri gelen uzmanlarının fikir ayrılığına düştüğü böylesine geçerli ihtilafların ötesinde, yanlış kanılar yüzünden ortaya çıkan sıkıcı sözde ihtilaflar da mevcuttur. Mesela, siz ve tartıştığınız kişi kelimeleri aynı anlamlara gelecek şekilde kullanmıyorsanız "yaşam", "zekâ" ya da "bilinç" üzerine tartışarak vakit harcamayın! Bu kitap Tablo 1.1'deki tanımları kullanıyor.
- Ayrıca Görsel 1.5'te bulunan sık rastlanan yanlış kanılara da dikkat: "2100 yılında süper zekâ kaçınılmaz olacak / imkânsızdır." "Yalnızca Ludditelar YZ konusunda endişeleniyor." "YZ'nin bilinçli ve/veya şeytani olması konusunda endişeler var ve önümüzde yalnızca yıllar kaldı." "Robotlar asıl endişe kaynağı." "YZ insanları kontrol edemez ve hedeflere sahip olamaz."
- Bölüm 2'den 6'ya, milyarlarca yıl önceki naçizane başlangıcından milyarlarca yıl sonraki olası kozmik geleceğine kadar zekânın hikâyesini keşfedeceğiz. İlk önce iş gücü, YZ silahları ve insan seviyesi YGZ arayışı gibi yakın dönem zorlukları irdeleyeceğiz, sonrasında da zeki makine ve/veya insanlarla mümkün olabilecek baş döndürücü gelecek olasılıklarını keşfedeceğiz. Acaba hangi seçenekleri tercih edeceksiniz merak ediyorum!
- Bölüm 7'den 9'a, sıkı bilimsel gerçekliklerden hedefler, bilinç ve anlamın keşfine geçecek ve istediğimiz geleceği yaratmak için şu anda neler yapabileceğimizi araştıracağız.
- YZ ile yaşamın geleceği hakkındaki bu tartışmayı zamanımızın en önemli tartışması olarak görüyorum. Lütfen bana katılın!

Bölüm 2:



Madde Zekâ Kazanıyor

Hidrojen... yeterince zaman verildiğinde insanlara dönüşür.

Edward Robert Harrison, 1995

Büyük Patlama'dan bugüne geçen 13.8 milyar yıl içerisinde gerçekleşen en şaşırtıcı gelişmelerden biri de aptal ve cansız maddenin zekâ kazanmasıdır. Bu nasıl olabildi ve şeyler gelecekte daha ne kadar zekileşebilir? Evrenimizde zekânın tarihi ve kaderi hakkında bilimin söyleyebileceği neler var? Bu soruların üstesinden gelebilmek adına, bu bölümü zekânın temellerine ve temel yapıtaşlarına ayıralım. Bir madde kütlesinin zekâya sahip olduğunu söylemek ne anlama gelir? Ya da bir objenin hatırlayabildiği, işlem yapabildiği ve öğrenebildiğini söylemek?

Zekâ Nedir?

Eşim ve ben geçtiğimiz günlerde İsveç Nobel Vakfı'nca düzenlenen bir yapay zekâ sempozyumuna katılma şansını elde ettik. Bir panelde ileri gelen YZ araştırmacılarından zekâyı tanımlamaları istendi; bir ortak görüşe varmadan uzun uzun tartıştılar. Bunu oldukça eğlenceli bulmuştuk: zeki zekâ araştırmacıları ara-

sında bile zekânın ne olduğuna dair bir uzlaşma yoktu! Demek ki zekânın ihtilafsız, "doğru" bir tanımı bulunmuyordu. Onun yerine, birbiriyle rekabet eden tanımlar mevcuttu. Mesela, mantık kapasitesi, anlama, planlama, duygusal bilgi, öz farkındalık, yaratıcılık, problem çözme ve öğrenme gibi...

Zekânın geleceğine dair yaptığımız keşiflerde, mümkün olan en geniş ve kapsayıcı bakışı takınmayı istiyoruz, yalnızca bugüne kadar ortaya çıkmış zekâlarla sınırlı kalmayı değil. Bu yüzden de son bölümde yaptığım tanımlama ve bu kitap boyunca kelimeyi kullanış biçimim oldukça geniş:

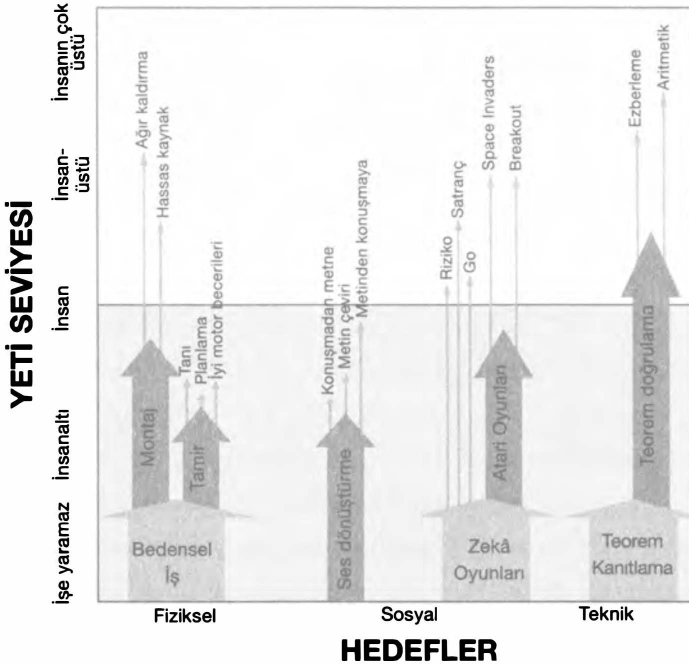
Zekâ = *karmaşık hedeflere ulaşabilme yetisi*

Anlama, öz farkındalık, problem çözme, öğrenme vb. hep karmaşık hedef örnekleri olduğu için, bu tanım yukarıda belirtilenlerin hepsini kapsayacak kadar geniştir. Ayrıca, bilgi ve yetileri uygulama gibi bir hedef de olabileceği için, *Oxford Sözlük*'teki tanımı "bilgi ve yeti edinme ve uygulama yeteneği" içerecek kadar da geniştir.

Pek çok olası hedef olduğu için, pek çok olası zekâ tipi vardır. Bizim tanımımızdan yola çıkarsak, insanları, insan olmayan hayvanları ya da makineleri IQ* gibi tek bir sayıyla ifade etmek pek de anlamlı olmuyor. Hangisi daha zekidir: Yalnızca satranç oynayabilen bir bilgisayar programı mı yoksa yalnızca Go oynayabilen mi? Bunun mantıklı bir cevabı yoktur çünkü ikisi de doğrudan kıyaslanamayacak farklı şeylerde iyidir. Ancak üçüncü bir program *tüm* hedeflere ulaşmada en azından onlar kadar iyi ve en az bir tanesinde açıkça daha iyiyse (mesela satrançta) diğer ikisinden daha iyidir diyebiliriz.

* Bunu daha iyi irdeleyebilmek için, olimpiyat seviyesindeki atletik başarılara ulaşabilme yetisinin "atletizm katsayısı" adı verilen, kısaltması da AQ olan, tek bir sayı ile tanımlanabileceği söylenseydi nasıl tepki vereceğinizi hayal edin. Böyle bir durumda tüm sporlarda altın madalyayı en yüksek AQ'ya sahip olimpiyat sporcusu kazanırdı.

Yetenek bir yelpazeye tekabül ettiğinden ve "ya hep ya hiç" bir özellik olmadığından dolayı da bir şeyin sınırda kalan vakalarda zeki olduğu ya da olmadığı konusunda konuşmak da pek anlamlı olmaz. Hangi insanlar konuşma hedefine ulaşabilme yetisine sahiptir? Yenidoğanlar mı? Hayır. Radyo spikerleri? Evet. Peki, yalnızca on kelime kurabilen bebeklere ne demeli? Ya da beş yüz kelime kurabilenler? Sınırı nereden çekeceğiz? Yukarıdaki tanımda belirsiz "karmaşık" kelimesini bilerek kullandım çünkü zekâ ve zeki olmayan arasına yapay bir çizgi çizmeyi denemek ilginç değil ve bu kelime, farklı hedeflere ulaşma yetisini basitçe derecelendirmek için çok daha faydalı.



Görsel 2.1: Karmaşık hedeflere ulaşabilme yetisi olarak tanımlanmış zekâ tek bir IQ ile değil, yalnızca tüm hedefleri kapsayan bir yeti yelpazesiyle ölçülebilir. Her bir ok bugünkü en iyi YZ sistemlerinin çeşitli hedeflere ulaşmada ne derece becerikli olduğunu gösterirken, her sistem yalnızca

belirli bir hedefi gerçekleştirebildiği için günümüz yapay zekâsının *dar* olma eğilimini de yansıtıyor. Öte yandan, insan zekâsı önemli biçimde geniştir: Sağlıklı bir çocuk neredeyse her şeyde kendini geliştirmeyi öğrenebilir.

Farklı zekâları bir sınıflandırmaya sokmak için bir diğer önemli ayrım da *dar* ve *geniş* zekâ arasındadır. IBM'in 1997 yılında satranç şampiyonu Garry Kasparov'u tahtından eden Deep Blue satranç bilgisayarı yalnızca oldukça dar bir görev olan satranç oynama işini yerine getirebiliyordu. Etkileyici yazılımı ve donanımına rağmen dört yaşındaki bir çocuğu bile üç taş oyununda alt edemezdi. Google DeepMind'ın DQN YZ sistemi görece biraz daha geniş görevleri yerine getirebilir: Bir dizi farklı eski Atari bilgisayar oyunlarını insan seviyesinde ya da daha iyi seviyede oynayabiliyor. Öte yandan, insan zekâsı çok daha geniş ve büyük bir yelpazede birçok beceride ustalaşabilme özelliğine sahip. Sağlıklı bir çocuğa yeterince alıştırma vakti verildiğinde yalnızca *herhangi bir* oyunda değil, herhangi bir dil, spor ya da meslekte oldukça iyi bir hale gelebilir. Bugün insanlar ve makinelerin zekâsını kıyasladığımızda, insanlar olarak genişlikte açık ara biz kazanıyoruz ancak makineler Görsel 2.1'de gösterildiği üzere küçük ama artan sayıdaki dar alanlarda bizi geride bırakıyorlar. YZ araştırmasının kutsal arayışı mümkün olduğunca geniş bir "genel YZ" (ya da daha iyi bilindiği şekliyle *yapay genel zekâ*, YGZ) inşa etmekti: Öğrenme de dâhil olmak üzere mümkün olan her hedefe ulaşabilme yetisine sahip bir makine. Bunu Bölüm 4'te ayrıntılı biçimde inceleyeceğiz. YGZ terimi, *insan seviyesinde* yapay genel zekâyı kastetmek için YZ araştırmacıları Shane Legg, Mark Gubrud ve Ben Goertzel tarafından popülerleştirildi. Yani, herhangi bir hedefe en az insanlar kadar iyi ulaşabilme yetisi.[1] Onların tanımlarıyla devam edeceğim, böylece açık bir biçimde

kısaltmayı değiştirmediğim sürece (mesela, "süper insan YGZ" yazmadığım sürece), YGZ'yi "insan seviyesi YGZ" anlamında kullanacağım.*

"Zekâ" kelimesinin olumlu çağrışımları olsa da bizim bu kelimeyi tamamıyla nötr kullandığımızı belirtmekte fayda var. Yalnızca, hedeflerin iyi ya da kötü olmasından bağımsız bir biçimde karmaşık hedeflere ulaşma yetisi olarak kullanıyoruz. Zeki bir insan insanlara yardımcı olmakta ya da onlara zarar vermekte çok iyi olabilir. Hedefler konusunu Bölüm 7'de irdeleyeceğiz. Hedeflerle ilgili olarak, kimin hedeflerinden bahsettiğimizi de netleştirmemiz gerek. Diyelim ki gelecekteki yeni robotik kişisel yardımcınızın kendine ait hiçbir hedefi yok ancak ne yapmasını isterseniz yapacak ve siz de ondan mükemmel İtalyan yemeğini yapmasını istediniz. Eğer internete girer ve İtalyan yemek tariflerini, en yakındaki süpermarkete nasıl gidildiğini, makarnanın nasıl süzüldüğünü araştırır ve sonrasında başarılı bir şekilde gerekli malzemeleri alıp lezzetli bir yemek hazırlarsa, asıl hedef sizin olsa bile onu muhtemelen zeki olarak değerlendireceksiniz. Aslında, istekte bulunduğunuz anda sizin hedefinizi benimsedi ve onu kendi içinde kasiyere para ödemekten Parmesan rendelemeye uzanan küçük hedefler hiyerarşisine böldü. Bu anlamda, zeki davranış doğrudan hedef edinmeyle ilişkilidir.

* Bazıları YGZ'nin eş anlamlısı olarak "insan seviyesi YZ" ya da "güçlü YZ" terimlerini kullanmayı tercih ediyor ama ikisi de sorunlu. Bir cep hesap makinesi bile dar anlamda insan seviyesinde bir YZ'dir. "Güçlü YZ" teriminin zıddı "zayıf YZ'dir" ancak Deep Blue, Watson ve AlphaGo gibi dar YZ sistemlerine "zayıf" demek biraz garip geliyor.



Görsel 2.2: Hans Moravec'in "insan kabiliyeti manzarası" illüstrasyonu. Burada yükseltiler bilgisayarlar için zorluğu, yükselen deniz seviyeleri de bilgisayarların yapabildiklerini temsil eder.

Görsel 2.1'de de olduğu gibi görevlerin zorluklarını insanların onları gerçekleştirirken zorlanmalarına göre sınıflandırmamız doğaldır. Ancak bu, bilgisayarlar için ne kadar zor olacağına dair yanlış yönlendiren bir sonuç ortaya çıkarabilir. 314.159 ile 271.828'i çarpmak bir fotoğraftaki arkadaşı tanımaktan daha zor gelebilir ancak bilgisayarlar daha ben doğmadan önce bile aritmetikte bizim önümüzdeydiler fakat insan seviyesi görüntü tanıma çok yakın bir tarihte mümkün olabildi. Düşük seviye duyu-motor görevlerinin inanılmaz büyüklükte sayısal kaynağa ihtiyaç duymasına rağmen kolay gözükmesi Moravec paradoksu olarak bilinir ve beynimizin böyle görevlere kişiselleşmiş donanımının büyük bir kısmını ayırarak –hatta beynimizin dörtte birinden fazlasını– kolay gösterebiliyor olmasıyla açıklanır.

Hans Moravec'in bu metaforunu çok seviyorum ve Görsel 2.2'de onu resimlendirdim:

> Bilgisayarlar evrensel makinelerdir, potansiyelleri sınırsız bir görev yelpazesi üzerinde eş biçimde uzanır. İnsan potansiyeli ise hayatta kalmak için önemli olan alanlarda güçlüdür ama buna uzak alanlarda zayıftır. Alçak kısımlarında "aritmetik" ve "alışkanlık edinme" gibi etiketler, dağ eteklerinde "teorem kanıtlama" ve "satranç oynama," yüksek dağ tepelerinde de "hareket yeteneği," "göz-el koordinasyonu" ve "sosyal etkileşim" gibi etiketler bulunan bir "insan kabiliyeti manzarası" tahayyül edin. İlerleyen bilgisayar performansı bu manzarayı yavaşça basan su gibidir. Yarım yüz yıl evvel, alçak alanları kaplamaya başlayarak insan hesaplayıcılarını ve yazmanları dışarı itti ancak pek çoğumuz hâlâ kuru alandaydık. Şimdi de sel eteklere vardı ve oradaki mevzilerimiz çekilmeyi gözden geçiriyor. Tepelerimizde güvende hissediyoruz ancak bu hızla bir yarım yüzyıl sonra onlar da su altında kalacak. Benim önerim o günler yaklaştıkça gemiler inşa etmek ve denizcilik hayatını benimsektir![2]

Bu satırların yazıldığı zamandan bu yana geçen on yıllarda, yazarın tahmin ettiği gibi, deniz seviyesi dur durak bilmeden küresel ısınma gibi yükseldi ve bazı etekler (satranç gibi) uzun süredir sular altında kalmış durumda. Bir sonraki adımın ne olduğu ve bizim ne yapmamız gerektiği de bu kitabın kalan kısmının konusu.

Deniz seviyesi yükselmeyi sürdürürken, bir gün son noktaya ulaşabilir ve dramatik bir değişikliği tetikleyebilir. Bu kritik deniz seviyesi makinelerin YZ tasarımı gerçekleştirebileceği noktaya denk düşüyor. Bu noktaya ulaşmadan önce, deniz seviyesi *insanların* makineleri iyileştirmesi yüzünden yükseliyordu; bundan sonra ise, artış makineleri iyileştiren *makineler* tarafından muhtemelen insanların yapabileceğinden çok daha hızlı bir şekilde gerçekleşerek

birden tüm karayı sular altında bırakabilecek. Bu da Bölüm 4'te keşfedeceğimiz çarpıcı ve tartışmalı *tekillik* kavramına denk düşer.

Bilgisayar biliminin öncü isimlerinden Alan Turing ünlü bir biçimde bir bilgisayarın belirli bir minimum operasyon dizisini gerçekleştirmesi durumunda, yeterli zaman ve bellek verilirse *herhangi* başka bir bilgisayarın yapabileceği herhangi bir şeyi yapmaya programlanabileceğini kanıtlamıştı. Bu kritik eşiği aşan makinelere *evrensel bilgisayarlar* (ya da Turing evrensel bilgisayarları) denir; bugünkü tüm akıllı telefonlar ve dizüstü bilgisayarlar bu anlamda evrenseldir. Benzer biçimde, YZ tasarımı için gerekli olan kritik zekâ eşiğinin *evrensel zekâ* eşiği olduğunu düşünmeyi tercih ediyorum. Yeterince zaman ve kaynak verildiği takdirde, kendisini *herhangi* başka akıllı bir varlığın yapabileceği kadar iyi biçimde herhangi bir hedefe ulaşabilecek duruma getirebilir. Örneğin, daha iyi sosyal beceriler, öngörme yetisi ve YZ tasarım kabiliyeti istediğine karar verirse, bunları edinir. Eğer bir robot şirketi kurmayı öğrenmeye karar verirse, bunu yapabilir. Diğer bir deyişle, evrensel zekâ Yaşam 3.0'a evrilme potansiyelini sahiptir.

Yapay zekâ araştırmacıları arasındaki genel görüş en nihayetinde zekânın et, kan ya da karbon atomlarıyla değil bilgi ve hesaplamayla ilgili olduğu yönündedir. Bu da makinelerin bir gün en azından bizim kadar zeki olmaması için hiçbir temel sebep olmadığı anlamına gelir.

Fakat fiziğin gösterdiği kadarıyla, temelde her şey etrafta dolaşan madde ve enerjiyse, bilgi ve hesaplama nedir? Bilgi ve hesaplama gibi soyut, fiziksel varlığı olmayan ve uhrevi bir şey nasıl fiziksel bir şeyin içinde vücut bulabilir? Özellikle de etrafta fizik kurallarına göre hareket edip duran bir avuç aptal parçacık nasıl bizim zeki diyebileceğimiz davranışlar sergileyebilir?

Eğer bu sorunun cevabının bariz olduğunu düşünüyor ve makinelerin bu yüzyılda insanlar kadar zeki olabileceğini akla

yatkın buluyorsanız –mesela YZ araştırmacısı iseniz– lütfen bu bölümün kalan kısmını atlayıp doğrudan Bölüm 3'e geçin. Kalanlara gelince, sonraki üç kısmı özellikle sizin için yazmış olduğumu bilmek sizi sevindirecektir.

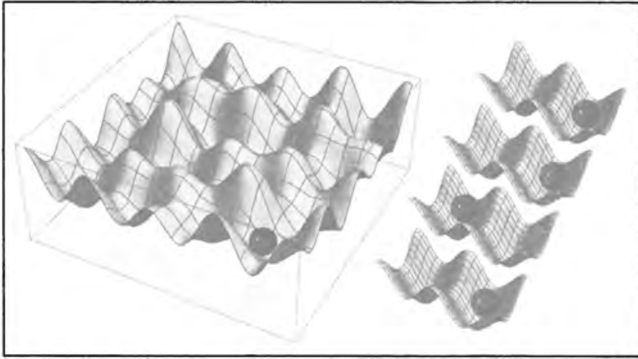
Bellek Nedir?

Eğer bir atlasın dünya hakkında *bilgi* içerdiğini söylersek, kitabın durumu ile (yani harf ve görsellere renklerini veren belli başlı moleküllerin konumları ile) dünyanın durumu (örneğin, kıtaların konumları) arasında bir ilişki olduğunu söylemiş oluruz. Eğer kıtalar başka yerlerde olsaydı, o moleküller de başka yerlerde olacaklardı. İnsanlar olarak bilgiyi depolamak için kitaplar ve beyinlerden disk sürücülere kadar pek çok farklı cihaz kullanırız ve hepsi şu özelliği paylaşır: Durumları önem verdiğimiz diğer şeylerin durumlarıyla ilişkilidir (böylece bize onlar hakkında bilgi verir).

Peki, hepsinde bulunan ve onları iyi bellek cihazları, yani bilgi depolama cihazları yapan temel fiziksel özellik nedir? Bunun cevabı hepsinin *pek çok farklı uzun ömürlü durumda olabilmeleridir*, yani gerekene dek bilgiyi kodlayacak kadar uzun ömürlü olmak. Basit bir örnek olarak, Görsel 2.3'teki gibi on altı farklı vadisi olan tepelik bir yüzeye bir top koyduğunuzu düşünün. Top bir kez yuvarlandı mı durur, on altı vadiden birindedir. Artık konumunu 1 ila 16 arasındaki bir sayıyı hatırlamak için kullanabilirsiniz.

Bu bellek cihazı görece sağlamdır çünkü birazcık sarsılsa ve dış güçler tarafından müdahale edilse bile, top onu koyduğunuz vadide kalma eğilimindedir, bu yüzden de halen daha hangi numara olduğunu söyleyebilirsiniz. Bu belleğin bu kadar sağlam olmasının sebebi topu vadisinden çıkarmak için gerekli olan enerjinin rastgele müdahalelerin verebileceğinden daha yüksek olmasıdır.

Aynı fikir hareket edebilen bir topa olduğundan daha da genel bir şekilde sağlam belleklere de uygulanabilir: Komplike bir fiziksel sistemin enerjisi her tür mekanik, kimyasal, elektrik ve manyetik özelliklere bağlıdır ve hatırlamasını istediğiniz durumdan sistemi uzaklaştırmaya yetecek kadar enerji almadığı sürece, durumu durağan olacaktır. Bu yüzden katıların durumları uzun ömürlüyken sıvılar ve gazların değildir. Eğer birinin adını altın bir yüzüğe kazırsanız, bu bilgi yıllar sonra bile orada kalacaktır çünkü altını yeniden şekillendirmek önemli ölçüde enerji gerektirir. Fakat bir göletin üzerine yazmaya çalışırsanız su yüzeyi birden değişeceği için saniyeler içinde kaybolacaktır.



Görsel 2.3: Fiziksel bir obje pek çok farklı sağlam durumda bulunabiliyorsa kullanışlı bir bellek cihazıdır. Soldaki top $2^4 = 16$ vadiden hangisinde bulunduğunu gösteren dört bitlik bilgiyi kodlayabilir. Birlikte, sağdaki dört top da her biri bir bit olmak üzere dört bit bilgi kodlar.

En basit hafıza cihazında yalnızca iki durağan durum vardır (Görsel 2.3, sağ). Bu sebeple bunu bir ikilik sayı (yani "bit") olarak kodlamayı düşünebiliriz: sıfır ya da bir. Daha komplike bellek cihazları tarafından depolanan bilgi de aynı şekilde çoklu bitlerle depolanabilir: Mesela, beraber ele alındığında, Görsel 2.3'te (sağ) gösterilen dört bit, 2 x 2 x 2 x 2 = 16 farklı durumda bulunabi-

lir, 0000, 0001, 0010, 0011, ..., 1111, böylece birlikte (soldaki) 16 durumlu daha karmaşık sistemle eşdeğer bellek kapasitesine sahip olurlar. Bitleri bilgi atomları olarak –bir araya geldiklerinde bilgiyi oluşturan, daha da küçültülemeyen bölünemeyen en küçük bilgi parçacıkları olarak– düşünebiliriz. Mesela, bilgisayarımda "word" kelimesini yazdım ve bilgisayarım onu belleğinde her biri 8 bit olan dört rakamlı 119 111 114 100 serisiyle temsil etti (her küçük harf, alfabesindeki sıra numarasının 96 fazlası olan bir sayıyla temsil edilir). *W* tuşuna bastığım anda bilgisayarım ekranında *w* harfini gösterdi. Bu görsel de bitlerle temsil edilir: 32 bit ekrandaki milyonlarca pikselin her birinin rengini belirler.

İki durumlu sistemleri üretmek ve onlarla çalışmak kolay olduğu için, pek çok modern bilgisayar bilgilerini bitlerle depolar ancak bu bitler bambaşka şekillerde vücut bulur. Bir DVD'de, her bit plastik yüzeyde verili her anda mikroskobik bir çukur olup olmadığına göre belirlenir. Disk sürücüde, her bit yüzeydeki bir noktanın iki şekilden hangisinde manyetize edildiğine göre şekillenir. Bilgisayarımın çalışan belleğinde, her bit çeşitli elektronların pozisyonlarına tekabül ederek, mikrokapasitör adı verilen bir cihazın yüklü olup olmayacağına karar verir. Bazı tür bitleri nakletmek de mümkündür, hem de ışık hızında. Mesela e-postanızı taşıyan optik fiberde, her bit verili bir anda lazer ışınının güçlü ya da zayıf olduğunu gösterir.

Mühendisler bitleri yalnızca dengeli ve okuması kolay olan (altın yüzük gibi) sistemlere değil, yazması kolay sistemlere de kodlamayı tercih ederler. Hard diskinizin durumunu değiştirmek altın bir yüzüğe isminizi kazımaktan daha kolaydır. Çalışması kolay ve seri üretimi ucuz olan sistemleri de tercih ederler. Fakat bunun yanında, bitlerin fiziksel objeler olarak nasıl temsil edildiklerini pek de önemsemezler – genelde siz de önemsemezsiniz çünkü bir önemi yoktur! Bir arkadaşınıza basılacak bir doküman

yollarsanız, bilgi hızlı bir silsile halinde hard diskinizdeki manyetizasyondan bilgisayarınızın çalışan belleğindeki elektrik yüklerine, kablosuz ağınızdaki radyo dalgalarına, modeminizdeki voltajlara, optik fiberdeki lazer atımlarına ve en son bir kâğıt parçasındaki moleküllere dönüşür. Diğer bir deyişle, *bilgi fiziksel materyalden bağımsız olarak kendi başına hayat bulur!* Gerçekten de bizim ilgilendiğimiz tek şey genelde bilginin bu maddeden bağımsız yanıdır: Eğer arkadaşınız gönderdiğiniz doküman hakkında konuşmak için sizi ararsa, muhtemelen voltajlar ve moleküllerden bahsetmeyecektir. İşte bu, zekâ gibi fiziksel varlığı olmayan bir şeyin somut fiziksel şeylerde vücut bulabileceğine dair ilk ipucumuz ve kısa süre içinde bu materyalden bağımsız olma fikrinin nasıl yalnızca bilgiyi değil hesaplama ve öğrenmeyi de kapsayabilecek şekilde çok daha derin olabileceğini de göreceğiz.

Materyalden bağımsızlık sebebiyle, zeki mühendisler yeni teknolojilere dayanarak, yazılımda hiçbir değişim gereksinimi olmadan, bilgisayarlarımızın içindeki bellek cihazlarını devamlı olarak çok daha iyileriyle değiştirebildiler. Sonuç Görsel 2.4'te görüldüğü üzere harikaydı: Son altmış yıl içinde, bilgisayar belleği her birkaç yılda bir yarı fiyatına inmişti. Sabit diskler yüz milyon kat ucuzlamıştı ve yalnızca depolamaktan ziyade hesaplamak için daha faydalı olan hızlı bellekler inanılmaz derecede on trilyon kat ucuzlamıştı. Alışveriş yaparken böyle "%99.9999999999 indirim" alabilseniz, New York'taki tüm binaları yaklaşık on sente ve şu ana kadar çıkarılmış tüm altını bir dolara alabilirdiniz.

Çoğumuz için, bellek teknolojisindeki harika gelişmelerin kişisel hikâyeleri de vardır. Lisedeyken 16 kilobayt belleği olan bir bilgisayar yarışmasına katılabilmek için bir şeker dükkânında çalıştığım günleri hatırlıyorum. Lisedeki sınıf arkadaşım Magnus Bodin'le o yarışma için bir kelime işlemcisi yapıp sattığımızda, kodu işleyeceği kelimeler için yeterince yer kalması için ultra

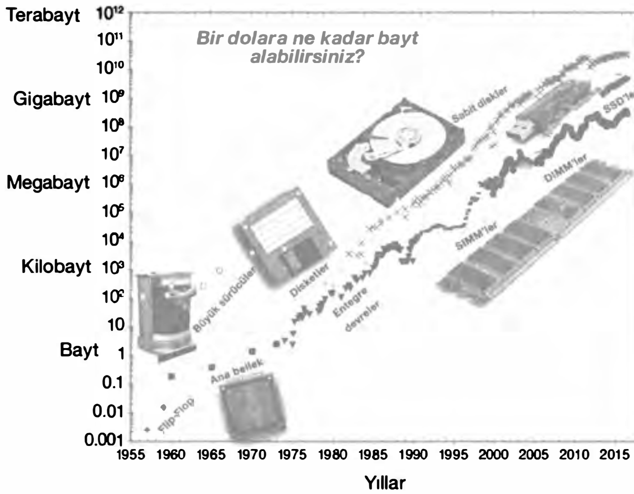
kompakt bir makine dilinde yazmamız gerekmişti. 70 kB depolayabilen disketlere alıştıktan sonra, 1,44 MB depolayabilen ve tüm bir kitabı içinde barındırabilen daha küçük 3,5 inçlik disketleri ve sonrasında da –şu anki şarkılardan yalnızca bir tanesini alabilecek- 10 MB depolayabilen ilk sabit diskimi gördüğümde ağzım açık kalmıştı. Ergenliğimdeki bu anılar daha dün üç bin kat fazla yere sahip bir sabit diske yüz dolar verdiğimi düşününce neredeyse gerçek dışı gibi geliyor.

Peki ya insanlar tarafından tasarlanan değil evrimleşen bellek cihazlarına ne demeli? Biyologlar nesiller arasında kendi şablonunu kopyalayan ilk yaşam formunun ne olduğunu henüz bilmiyorlar ancak muhtemelen çok ufaktı. Cambridge Üniversitesi'nden Philipp Holliger'in başını çektiği bir ekip, 2016 yılında 412 bit üretken bilgi kodlayan ve kendinden uzun RNA zincirleri kopyalayabilen bir RNA molekülü yaparak dünyadaki ilk yaşamın kısa, kendini kopyalayan RNA parçalarından oluştuğunu söyleyen "RNA dünyası" hipotezini desteklemişti. Şu ana kadar bilinen evrimleşmiş ve doğada kullanılmış en küçük bellek cihazı, *Candidatus Carsonella ruddii* bakterisinin 40 kilobayt depolayabilen genomudur. Bunun yanında, bizim insan DNA'mız bir filmle kıyaslanabilecek kadar, yaklaşık 1,6 gigabayt depolayabilmektedir. Son bölümde de dediğimiz gibi, beyinlerimiz genlerimizden çok daha fazla bilgi depolayabilir: elektriksel olarak (verili bir zamanda yüz milyar nöronunuzun hangilerinin ateşlendiğine göre) yaklaşık on gigabayt ve (sinapslarınızın farklı nöronları ne kadar güçlü bağladığına göre) kimyasal/biyolojik olarak yüz terabayt. Bu rakamları makine bellekleriyle kıyaslamak dünyanın en iyi bilgisayarlarının şu anda herhangi bir biyolojik sistemden daha yüksek belleğe sahip olduğunu gösteriyor, hızla düşen ve 2016'da birkaç bin dolar olan bir ücrete hem de.

Beyninizdeki bellek, bilgisayar belleğinden çok daha farklı çalışır. Yalnızca yapısı anlamında değil, nasıl kullanıldığı anlamında da. Bilgisayardan ya da sabit diskten anılarınızı *nerede* depolandığını söyleyerek çağırırken, beyninizden anıları *neyin* depolandığını belirleyerek çağırırsınız. Bilgisayar hafızanızdaki her bit grubunun nümerik bir adresi vardır ve bu bilgiyi çağırmak için bilgisayar hangi adrese bakması gerektiğini belirler. Tıpkı size, "Kitaplığıma git, üst raftan sağdan beşinci kitabı al ve sayfa 314'te ne yazdığını bana söyle," demem gibi. Aksine, beyninizden bilgiyi bir arama motorundan bulur gibi bulursunuz: önce bir bilgi parçası ya da onunla ilgili bir şey belirlersiniz ve ortaya çıkar. Eğer size, "Olmak ya da," desem veya bunu Google'da aratsam, "Olmak ya da olmamak, işte bütün mesele bu," cümlesinin çıkma ihtimali çok büyüktür. Hatta bu alıntının başka bir kısmını kullansam veya bir şekilde karmaşıklaştırsam da sonuç aynı olur. Böyle bellek sistemlerine *öz çağrışımsal* denir çünkü adresle değil çağrışımla hatırlarlar.

1982'de yazdığı ünlü bir makalede, fizikçi John Hopfield bağlantılı nöronlardan oluşan bir ağın özçağrışımsal bir bellek olarak nasıl işlediğini göstermişti. Bu temel fikri çok güzel bulurum, dahası çoklu kararlı duruma sahip herhangi bir fiziksel sistem için de işler. Mesela, iki vadisi olan bir yüzeydeki bir topu düşünelim, tıpkı Görsel 2.3'te bulunan tek bitli sistem gibi. Bu yüzeyi öyle şekillendirelim ki topun durduğu iki minimum noktanın x koordinatları, sırasıyla $x = \sqrt{2} \approx 1.41421$ ve $x = \pi \approx 3.14159$ olsun. Eğer π'nin yalnızca 3'e yakın olduğunu hatırlarsanız, topu x=3'e koyarsınız ve en yakın minimum noktaya yuvarlanırken daha kesin bir π-değeri ortaya çıkarışını izlersiniz. Hopfield, nöronların karmaşık ağının sistemin yerleşebileceği pek çok enerji-minimum noktalarıyla paralel bir çevre sunduğunu fark etti ve sonrasında,

büyük bir karmaşaya mahal vermeden her bin nöron için yüz otuz sekiz kadar anının sıkıştırılabileceğini kanıtlandı.



Görsel 2.4: Son altmış yılda bilgisayar belleği her birkaç yılda bir yarı fiyatına inerek, her yirmi yılda bin kat ucuzlamış oldu. Bir bayt sekiz bite denktir. John McCallum'un izniyle, http://www.jcmit.net/memoryprice.htm.

Hesaplama Nedir?

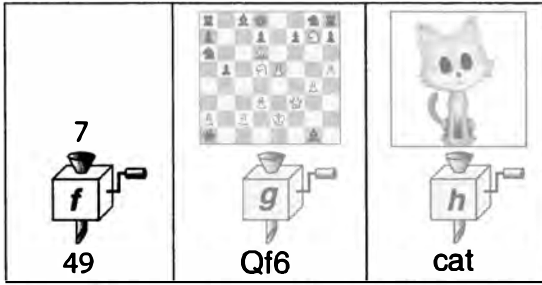
Fiziksel bir objenin nasıl bilgi hatırlayabildiğini gördük. Peki, nasıl hesaplıyor?

Hesaplama bir bellek durumunun bir diğerine dönüşümüdür. Diğer bir deyişle hesaplama, matematikçilerin *fonksiyon* dediği şeyi uygulayarak, bilgiyi alır ve dönüştürür. Ben fonksiyonu bilgi için bir kıyma makinesi olarak görüyorum. Görsel 2.5'te de görüldüğü gibi: Yukarıdan bilgiyi koyuyorsunuz, kolu çeviriyorsunuz ve aşağıdan işlenmiş bilgiyi alıyorsunuz. Hem de bunu istediğiniz kadar farklı girdilerle tekrarlayabiliyorsunuz. Bu bilgi

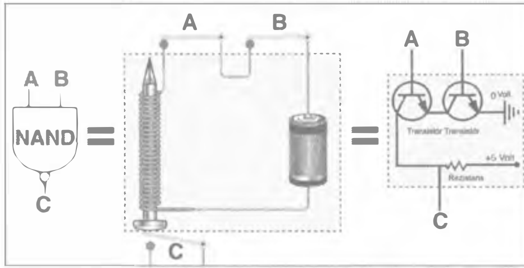
işlem, aynı girdiyle tekrarladığınızda her seferinde aynı çıktıyı vermesi bakımından deterministtir de.

Çok basit gözükse de bu fonksiyon fikri inanılmaz derecede geneldir. Bazı fonksiyonlar çok daha önemsizdir. Mesela NOT (DEĞİL) adı verilen ve tek bir bit girdi alıp tersini çıktı olarak veren, yani biri sıfır, sıfırı da bir yapan fonksiyon gibi. Okulda öğrendiğimiz fonksiyonlar hesap makinesindeki tuşlara denk düşer; bir ya da birçok numara girdisi alır ve tek bir sayı çıktısı verir – mesela x^2 fonksiyonu bir sayı girdisi alır ve onun kendisiyle çarpılmış halinin çıktısını verir. Diğer fonksiyonlar ise ileri derece karmaşık olabilirler. Örneğin, rastgele bir satranç pozisyonunu temsil eden bitleri girdi olarak alacak ve yapılabilecek en iyi hamlenin çıktısını verecek bir fonksiyonunuz varsa, bunu Dünya Bilgisayar Satranç Şampiyonu'nu yenmekte kullanabilirsiniz. Eğer dünyanın tüm finansal datasını girdi olarak alan ve alınacak en iyi hisselerin çıktısını veren bir fonksiyona sahipseniz, kısa zamanda aşırı derecede zengin olursunuz. Pek çok YZ araştırmacısı kariyerlerini belli başlı fonksiyonları nasıl kullanabileceklerini bulmaya adamıştır. Mesela, makine çevirisi araştırmasının hedefi bir dildeki bir metni temsil eden bitleri girdi olarak alacak ve aynı metni başka bir dilde temsil eden bitleri çıktı olarak verecek bir fonksiyon geliştirmektir. Otomatik resim yazısı araştırmasının hedefi ise bir görseli temsil eden bitleri girdi olarak alıp onu tasvir eden metni temsil eden bitleri çıktı olarak vermektir (Görsel 2.5, sağ).

Diğer bir deyişle, eğer yüksek oranda karmaşık fonksiyonları uygulayabilirseniz, yüksek oranda karmaşık hedefleri yerine getirebilen zeki bir makine inşa edebilirsiniz. Bu da maddenin nasıl zeki olabileceğine dair sorumuzu tam odağa taşıyor: Yani görünüşte zekâya sahip olmayan bir madde bütünü nasıl komplike bir fonksiyonu hesaplayabiliyor?



Görsel 2.5: Hesaplama, matematikçilerin *fonksiyon* dediği şeyi uygulayarak, bilgiyi alır ve dönüştürür. Fonksiyon *f* (solda) bir sayıyı temsil eden bitleri alır ve onun karesini hesaplar. Fonksiyon *g* (ortada) bir satranç pozisyonunu temsil eden bitleri alır ve beyaz için en iyi hamleyi hesaplar. Fonksiyon *h* (sağda) bir görseli temsil eden bitleri alır ve onu tasvir eden metni hesaplar.



Görsel 2.6: NAND kapısı A ve B bitlerini girdiler olarak alır ve A=B=1 ise C=0, değilse C=1 kuralına göre, C bitini çıktı olarak verir. Pek çok fiziksel sistem NAND kapısı olarak kullanılabilir. Ortadaki örnekte, anahtarlar 0=açık ve 1=kapalı olacak şekilde bitler olarak yorumlanabilir ve A ile B anahtarlarının ikisi de kapalıyken bir elektromıknatıs C anahtarını açar. En sağdaki örnekte, voltajlar (elektriksel potansiyel) 1=5 volt ve 0=0 volt olacak şekilde ayarlanmıştır, A ve B kablolarının ikisi de beş volttayken iki transistör elektrik iletir ve C kablosu yaklaşık sıfır volta düşer.

Altın bir yüzük ya da diğer statik bellek cihazları gibi hareketsiz kalmak yerine, karmaşık *dinamikler* sergilemelidir ki gelecekteki durumu komplike (ve kontrol edilebilir/programlanabilir) bir biçimde şimdiki durumuna bağlı olsun. Atom düzenlemesi ilgi çekici bir değişimin olmayacağı sert bir katıdan daha az düzenli

olmalı ama sıvı ya da gazdan daha düzenli olmalıdır. Özel olarak da sistemin şöyle bir özelliğe sahip olmasını isteriz: Girdi bilgisini kodlayacağı bir duruma soktuğumuzda, bir süre fizik kurallarına göre evrilmeli ve sonrasında da ortaya çıkan son durumu çıktı bilgisi olarak yorumlamalı. İşte o zaman çıktı, girdinin arzu edilen bir fonksiyonu olur. Durum buysa, sistemimizin fonksiyonumuzu hesapladığını söyleyebiliriz.

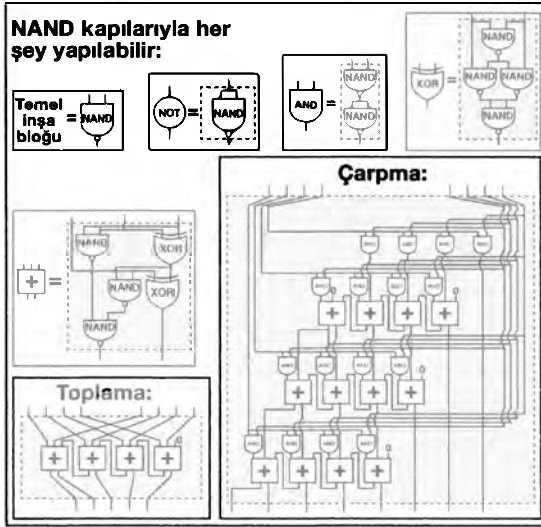
Bu fikrin ilk örneği olarak, *NAND kapısı** adı verilen basit (ama oldukça önemli) bir fonksiyonu zeki olmayan basit maddelerden nasıl yapabileceğimizi öğrenelim. Bu fonksiyon iki bit girdi alır ve bir bit çıktı verir: İki girdi de 1 ise, 0 çıktı verir; diğer tüm durumlarda çıktısı 1'dir. Eğer bir pil ve elektromıknatısla iki anahtarı seri bağlarsak, elektromıknatıs ilk anahtar *ve* ikinci anahtar bağlantısı "açık" hale getirilirse açık olur. Elektromıknatısın altına, Görsel 2.6'daki gibi üçüncü bir anahtar koyalım. Mıknatıs açıldığında bu anahtarı çekerek açabilsin. Eğer ilk iki anahtarı girdi bitleri ve üçüncüyü de çıktı biti olarak alırsak (0=anahtar açık, 1=anahtar kapalı), elimizde bir NAND kapısı olur: Üçüncü anahtar yalnızca ilk ikisi kapalıysa açık olur. Daha pratik NAND kapıları yapmak için çok başka yollar da vardır. Mesela, Görsel 2.6'da (sağda) gösterildiği gibi transistörler kullanılabilir. Bugünün bilgisayarlarında, NAND kapıları silikon plakalara otomatik olarak işleyen mikroskobik transistörler ve diğer bileşenlerden yapılır.

Bilgisayar biliminde NAND kapılarının *evrensel* olduğunu söyleyen muazzam bir teorem vardır, yani NAND kapılarını birbirlerine bağlayarak *herhangi bir* iyi tanımlanmış fonksiyonu uygulayabilirsiniz.** Yani eğer yeterince NAND kapısı yapabi-

* NAND, NOT AND'in (VE DEĞİL) kısaltmasıdır: Bir AND (VE) kapısı ilk girdi ve ikinci girdi 1 ise çıktı olarak 1 verir, NAND kapısı bunun tam tersidir.

** "İyi tanımlanmış fonksiyonu" matematikçiler ve bilgisayar bilimcilerin "hesaplanabilir fonksiyon" dedikleri şey anlamında kullanıyorum. Yani sınırsız bellek ve zaman ile varsayımsal bir bilgisayar tarafından hesaplanabilen bir fonksiyon olarak. Alan Turing ve Alonzo Church tanımlanabilen ama hesaplanamayan fonksiyonların da var olduğunu kanıtladılar.

lirseniz, her şeyi hesaplayabilen bir cihaz da yapabilirsiniz. Eğer bunun nasıl işlediğine dair biraz daha bilgi isterseniz, yalnızca NAND kapıları kullanarak sayıları nasıl çarpabileceğimizi gösteren bir şemayı Görsel 2.7'ye koydum.



Görsel 2.7: *Herhangi bir* iyi tanımlanmış hesaplama yalnızca NAND kapılarını doğru biçimde birleştirerek yapılabilir. Mesela, yukarıdaki toplama ve çarpma modülleri 4 bit ile temsil edilen iki ikilik sayıyı girdi olarak alır ve sırasıyla 5 ve 8 bit ile temsil edilen bir ikilik sayıyı çıktı olarak verir. Küçük modüller NOT, AND, XOR ve + (üç ayrı biti 2 bitlik ikilik bir sayıyla toplayan fonksiyon) da NAND kapılarından yapılır. Bu şekli tamamıyla anlamak oldukça zordur ve kitabın devamını takip etmek için tamamen gereksizdir; bunu buraya yalnızca evrensellik fikrini göstermek için ekliyorum. Bir de içimdeki ineği tatmin etmek için.

MIT araştırmacıları Norman Margolus ve Tommaso Toffoli rastgele hesaplamalar yapabilen tüm maddelere *computronium* ismini verdiler. Az önce bir *computronium* yapmanın o kadar da zor olmadığını gördük: Maddenin yalnızca istenen şekilde birleştirilmiş NAND kapılarını uygulayabiliyor olması yeterli.

Gerçekten de çok sayıda *computronium* çeşitleri var. Başka bir versiyonu da NAND kapılarını iki girdi de 0 iken çıktısı 1 olan NOR kapılarıyla değiştirerek yapılabilir. Bir sonraki kısımda, *computronium* gibi davranış gösterebilen, rastgele hesaplamaları uygulamaya koyabilen sinirsel ağları göreceğiz. Bilim insanı ve girişimci Stephen Wolfram komşu bitlerin yaptıklarına göre tekrarlı bir biçimde bitlerini güncelleyen hücresel otomasyon isimli basit cihazlarda da aynı şeyin olduğunu göstermişti. 1936 yılında bile, bilgisayar öncüsü Alan Turing dönüm noktası olan bir makalesinde bir bant şeridindeki sembolleri manipüle edebilen basit bir makinenin (şu anda "evrensel Turing makinesi" olarak biliniyor) rastgele hesaplamaları da uygulayabileceğini kanıtlamıştı. Özet olarak, maddenin yalnızca iyi tanımlanmış hesaplamaları yapması mümkün olmakla kalmaz, bunu bir dizi farklı şekilde yapması da olasıdır.

Daha önceden bahsettiğimiz üzere, Turing 1936 yılında yayımladığı makalesinde bundan çok daha derin bir şeyleri de kanıtlamıştı: bir bilgisayar belirli bir miktarda asgari operasyonu gerçekleştirebiliyorsa, yeterli kaynak sağlandığında *evrenseldir*, yani diğer bilgisayarların yaptığı her şeyi de yapabilir. Turing, makinesinin evrensel olduğunu gösterdi ve bunu tekrar fiziğe bağlarsak, bu evrensel bilgisayarlar ailesinin NAND kapıları ağı ve bağlantılı nöronlar ağı kadar çeşitli nesneleri de içerdiğini zaten görmüştük. Gerçekten de Stephen Wolfram, hava sistemlerinden beyinlere, basit olmayan fiziksel sistemlerden *çoğunun* rastgele büyük ve uzun süreli yapılabilirse evrensel bilgisayarlar olacağını savunuyordu.

Tam olarak aynı hesaplamanın *herhangi bir* evrensel bilgisayar üzerinde yapılabileceği gerçeği, tıpkı bilgi gibi *hesaplamanın da maddeden bağımsız* olduğu anlamına gelir: Fiziksel maddeden bağımsız olarak, kendi başına bir hayat sürebilir! Yani eğer gelecek-

teki bir bilgisayar oyununda bilinçli süper zeki bir karakterseniz, Windows masaüstünde mi, bir Mac OS laptopta mı yoksa bir Android telefonda mı olduğunuzu bilmenizin yolu olmaz çünkü maddeden bağımsız olurdunuz. Ayrıca mikroişlemcinin kullandığı transistörlerin ne tür olduğunu da bilme ihtimaliniz olmazdı.

Aslında fizikte çok fazla güzel örneği olduğu için bu maddeden bağımsız olma fikrini takdir etmeye başladım. Mesela dalgalar: Hız, dalga boyu ve frekans gibi özellikleri vardır ve biz fizikçiler hangi maddede dalga olduklarını bilmeye hiçbir ihtiyaç duymadan uydukları denklemleri çalışırız. Bir şey duyduğunuzda, hava dediğimiz gaz karışımında moleküllerin zıplamasından ortaya çıkan ses dalgalarını tespit edersiniz ve bu dalgalar üzerine her tür farklı ilginç şeyi –yoğunluğunun uzaklığın karesine göre nasıl azaldığını, açık kapılardan geçerken nasıl büküldüklerini ve duvarlardan sekerek nasıl eko yaptıklarını– havanın neden yapılmış olduğunu bilmeden hesaplayabiliriz. Aslında moleküllerden yapılmış olduğunu bile bilmemize gerek yok: Oksijen, azot, karbondioksit ve benzeri hakkında tüm detayları unutabiliriz çünkü dalganın bileşeninin ünlü dalga denklemine giren ve önemli olan tek özelliği, ölçebileceğimiz bir sayıdır: dalganın hızı, bu da hava için, yaklaşık saniyede üç yüz metredir. Gerçekten de geçen bahar bir derste MIT öğrencilerime anlattığım bu dalga denklemi, atomlar ve moleküllerin varlığı fizikçiler tarafından tespit edilmeden çok önce keşfedilmiş ve kullanıma sokulmuştu!

Bu dalga örneği üç önemli noktanın altını çiziyor. İlki, maddeden bağımsız olmak maddenin gereksiz olduğu anlamına gelmez, daha çok ayrıntılarının büyük bir kısmının önemli olmadığını söyler. Eğer gaz olmazsa bir gazda ses dalgalarına sahip olamayacağınız açık ama herhangi bir gaz yeterli olur. Benzer bir şekilde, madde olmadan hesaplama da olmaz ama NAND kapıları, bağlantılı nöronlar ya da evrensel hesaplamaya izin verecek başka tür yapı

blokları şeklinde düzenlenebildiği ölçüde herhangi bir madde işi görür. İkincisi, maddeden bağımsız olma durumu kendi başına bir hayata, maddesinden bağımsız olarak sahip olabilir. Bir dalga göl boyunca, su moleküllerinden hiçbiri hareket etmese de gidebilir – statta Meksika dalgası yapan taraftarlar gibi aşağı yukarı gidip gelirler sadece. Üçüncüsü, genelde ilgilendiğimiz tek şey maddeden bağımsız olma özelliğidir. Bir sörfçü genelde dalganın konumu ve yüksekliğiyle, moleküler dizilimine nazaran daha çok ilgilenecektir. Bunun bilgi için doğru olduğunu görmüştük, hesaplama için de doğru: Eğer iki programcı birlikte kodlarında bir hatayı bulmaya çalışıyorlarsa, muhtemelen transistörlerden bahsetmeyeceklerdir.

Şu anda, somut fiziksel şeylerin zekâ gibi manevi, soyut ve uhrevi bir şeyi hissedebilen bir başka şeye nasıl mahal verebileceğine dair sorduğumuz açılış sorumuza bir cevap bulduk: Fiziksel değilmiş gibi geliyor çünkü maddeden bağımsız, fiziksel detaylara bağlı olmayan ya da onları yansıtmayan bir hayatı kendi başına sürebilir. Kısacası hesaplama, parçacıkların uzay zaman düzeninde bir örüntüdür ve önemli olan parçacıklar değil örüntünün ta kendisidir! Maddenin artık bir önemi yoktur.

Diğer bir deyişle, donanım maddedir ve yazılım da örüntüdür. Hesaplamanın maddeden bağımsız olması YZ'nin mümkün olabileceğini ima eder: Zekânın et, kan ya da karbon atomlarına ihtiyacı yoktur.

Maddeden bağımsız olabildiği için, kurnaz mühendisler yazılım değiştirmeden bilgisayarlarımızın içindeki teknolojileri defalarca daha iyileriyle değiştirebildiler. Sonuçlar bellek cihazlarında olduğu kadar harikaydı. Görsel 2.8'de görebileceğiniz üzere, hesaplama her birkaç yılda bir yaklaşık yarısı kadar ucuz olmaya başlıyor ve bu trend bir yüzyıldır, ninelerimin doğduğu ana göre bilgisayar masraflarını milyon kere milyon kere milyon (10^{18}) kat daha ucuz yaparak devam ediyor. Her şey milyon kere milyon kere

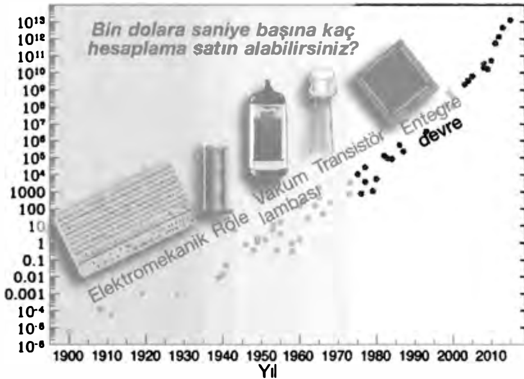
milyon kez ucuzlasaydı, bir sentin yüzde biriyle dünyada bu sene üretilen tüm ürün ve hizmetleri alabilirdiniz. Maliyetlerdeki bu dramatik düşüş elbette hesaplamanın bugünlerde neden her yerde olduğunu, geçmişin bina büyüklüğünde hesaplama tesislerinden evlerimize, arabalarımıza ve ceplerimize –hatta bazen ayakkabılar gibi beklenmedik yerlere bile– sıçradığını da açıklıyor.

Teknolojimiz niçin düzenli aralıklarla gücünü ikiye katlayarak matematikçilerin katsal büyüme dediği davranışı sergiliyor? Gerçekten, neden bu gelişme yalnızca transistörlerin minyatürleşmesi (*Moore yasası* olarak bilinen bir trend) anlamında değil de daha geniş bir anlamda bir bütün olarak hesaplama (Görsel 2.8), bellek (Görsel 2.4) ve genom diziliminden beyin görüntülemeye kadar bir dizi başka teknoloji için de gerçekleşiyor? Ray Kurzweil bu devamlı ikiye katlama durumuna "ivmelenen getiriler kanunu" diyor.

Devamlı ikiye katlamanın doğada bulunduğunu bildiğim tüm örneklerinde de aynı temel sebep mevcut ve bu teknolojik olanı da istisna değil: her adım bir sonrakini yaratıyor. Mesela siz ana rahmine düşmenizden hemen sonra bir katsal büyüme geçirdiniz: hücrelerinizin her biri bölündü ve günlük yaklaşık iki hücre ortaya çıkararak toplam hücre sayınızın her gün 1, 2, 4, 8, 16 diye artmasını sağladı. *Enflasyon* olarak bilinen, kozmik köklerimize dair en popüler bilimsel teoriye göre, bebek evrenimiz tıpkı sizin gibi katsal olarak büyüdü, bir atomdan daha küçük ve hafif bir zerrecik teleskoplarımızla gördüğümüz tüm galaksilerden daha kocaman olana dek, devamlı olarak boyutunu düzenli aralıklarla ikiye katladı. Sebep her ikiye katlama adımının bir sonrakini ortaya çıkardığı bir süreçti. Bu, teknolojinin ilerlemesine de benzer: Bir teknoloji iki kat güçlü olduğunda iki katı kadar güçlü olacak bir teknolojiyi tasarlayıp inşa etmekte

kullanılabilir, bu da Moore yasasında olduğu gibi tekrar eden bir ikiye katlama yetisini ortaya çıkarır.

Teknolojik gücümüzün ikiye katlanması kadar düzenli bir şekilde gerçekleşen bir başka şey de ikiye katlamanın sona erdiğine dair iddiaların ortaya çıkmasıdır. Evet, Moore yasası elbette sona erecek, yani transistörlerin küçülmesi için fiziksel bir sınır var. Ancak bazı insanlar Moore yasasının teknolojik gücümüzün devamlı ikiye katlanmasıyla eş anlamlı olduğunu yanlış biçimde varsayıyorlar. Tam aksine, Ray Kurzweil, Moore yasasının, Görsel 2.8'de gösterildiği gibi, hesaplamada katsal büyümeyi getiren ilk değil beşinci teknolojik paradigma olduğunu söyler: bir teknoloji gelişmeyi durdurduğunda onu daha da iyisiyle değiştiririz. Vakum lambalarını artık küçültemeyince, onları elektronların iki boyutta hareket edebildiği transistör ve entegre devrelerle değiştiririz. Bu teknoloji sınırlarına ulaşınca, deneyebileceğimiz daha pek çok alternatif olacak; mesela, üç boyutlu devreleri ve yapacağımızı yapmak için elektronlardan başka şeyleri kullanmak gibi.



Görsel 2.8: 1900'den beri, hesaplama her birkaç yılda bir yarı fiyatına düşmüştür. Bu grafik, hesaplama gücünü bin dolara satın alınabilecek saniye başına yüzer nokta işlemleri (FLOPS) açısından göstermektedir.[3] Yüzer nokta operasyonunu tanımlayan özel hesaplama bit döndürme ya da NAND ölçümü gibi yaklaşık 10^5 temel mantıksal operasyona tekabül eder.

Kimse bir sonraki çığır açıcı hesaplama maddesinin ne olacağını bilemez ancak fizik kurallarının bize dayattığı sınırların kesinlikle yakınında olmadığımızı biliyoruz. MIT'den arkadaşım Seth Lloyd bu temel sınırın ne olduğunu bulmaya çalıştı ve Bölüm 6'da çok daha detaylı bir biçimde inceleyeceğimiz gibi bu limit, bir madde bütününün yapabileceği hesaplama yönünden, bugünün en ileri teknolojisinin 10^{33} katıdır. Yani bilgisayarlarımızın hesaplama gücünü her birkaç yılda bir ikiye katlamayı başarsak bile, sınıra ulaşmamız iki yüz yıldan fazla sürecek.

Tüm evrensel bilgisayarlar aynı hesaplamaları yapabiliyor olsalar da bazıları diğerlerinden daha verimli çalışır. Mesela, milyonlarca çarpma gerektiren bir hesaplama, Görsel 2.6'daki gibi ayrı transistörlerden inşa edilmiş ayrı çarpma modüllerinden milyonlarcasını gerektirmez: Yalnızca bir tane böyle modül yeterlidir çünkü bu modülü gerekli girdilerle pek çok kez üst üste kullanabilir. Bu verimlilik ruhuyla, çoğu modern bilgisayar hesaplamaların çoklu zaman adımlarına bölündüğü bir paradigma kullanır. Bu süreç boyunca bilgi, bellek modülleri ve hesaplama modülleri arasında gidip gelir. Bu hesaplama mimarisi 1935 ile 1945 arasında bilgisayarın öncüleri olan Alan Turing, Konrad Zuse, Presper Eckert, John Mauchly ve John von Neumann tarafından geliştirildi. Daha da özel olarak, bilgisayar belleği hem veri hem de yazılımı (bir program, yani veriyle ne yapılacağına dair bir talimatlar listesi) depolar. Her zaman adımında, bir merkezî işlemci ünitesi (CPU) programda bir sonraki talimatı yürütür. Bu program bazı basit fonksiyonların verinin belli bir kısmına uygulanmasını sağlar. Bilgisayarın bir sonraki adımda ne yapılacağının hesabını tutan kısmı, *program sayacı* adı verilen, programdaki mevcut satır sayısını depolayan, belleğin bir başka kısmıdır. Bir sonraki talimata gitmek için, program sayacına yalnızca (1) ekler. Başka bir program satırına atlamak için, o satır numarasını program sayacına kopyalamak yeterlidir; "if (eğer)" komutları ve döngüler de bu şekilde işler.

Bugünün bilgisayarları, modüllerin bu yeniden kullanımının bir kısmını zeki biçimde ortadan kaldıran *paralel işleme* sayesinde ekstra hız kazanır: Eğer bir hesaplama paralel yapılabilen parçalara ayrılabiliyorsa (çünkü bir parçanın girdisi diğerinin çıktısına ihtiyaç duymuyordur), bu durumda parçalar donanımın farklı kısımlarınca eş zamanlı hesaplanabilir.

En üst düzey paralel bilgisayar *kuantum bilgisayarı*dır. Kuantum hesaplama öncüsü David Deutsch tartışmalı bir biçimde "kuantum bilgisayarların çoklu evrende kendilerinin çok sayıda versiyonuyla bilgi paylaştıklarını" ve bu diğer versiyonlardan yardım alarak bizim evrenimizde cevapları daha hızlı alabildiğini ileri sürmüştür.[4] Henüz önümüzdeki on yıllarda ticari olarak rekabetçi bir kuantum bilgisayarı yapılabilir mi bilmiyoruz çünkü hem kuantum fiziğinin bizim düşündüğümüz gibi çalışıp çalışmadığına hem de göz korkutucu teknik zorlukların üstesinden gelme yetimize bağlı. Ancak dünyanın dört bir yanından şirketler ve hükümetler bu ihtimale yılda on milyonlarca dolar yatırıyor. Kuantum bilgisayarları olağan hesaplamaların hızını artıramıyor olsa da şifre sistemlerini kırmak ve sinirsel ağları eğitmek gibi özel hesaplama türlerini hızlandıracak zekice algoritmalar geliştirildi. Bir kuantum bilgisayarı atomlar, moleküller ve yeni materyaller de dâhil olmak üzere kimya laboratuvarlarındaki ölçümlerin yerini tıpkı rüzgâr tünellerindeki ölçümlerin yerini geleneksel bilgisayarlardaki simülasyonların alması gibi alarak kuantum mekanik sistemlerin davranışını da etkili bir biçimde taklit edebiliyor.

Öğrenmek Nedir?

Bir hesap makinesi beni aritmetik yarışmasında harcayabilecek olsa da hızını ya da doğruluğunu ne kadar pratik yaparsa yapsın geliştiremez. Öğrenemez: Mesela, karekök tuşuna her bastığımda

tam olarak aynı fonksiyonu tam olarak aynı şekilde hesaplar. Benzer bir biçimde, beni satrançta yenen ilk bilgisayar programı da hatalarından asla ders almaz ancak zeki programcısının bir sonraki iyi hamleyi hesaplaması için tasarladığı fonksiyonu uygular. Öte yandan, Magnus Carlsen beş yaşında ilk satranç maçını kaybettiğinde, on sekiz yıl sonra onu Dünya Satranç Şampiyonu yapacak bir öğrenme sürecine başlamıştı.

Öğrenme yetisi genel zekânın en ilgi çekici yönlerinden biri olabilir. Görünüşte zekâya sahip olmayan bir maddenin hatırlayıp hesap yapabildiğini görmüştük ancak nasıl öğrenebilir? Zor bir soruya cevap bulmanın bir fonksiyonu hesaplamaya denk düştüğünü ve doğru şekilde düzenlenmiş maddenin herhangi bir hesaplanabilir fonksiyonu hesaplayabileceğini görmüştük. İnsanlar hesap makinelerini ve satranç programlarını ilk kez oluşturduğunda, düzenlemeyi *biz* yapmıştık. Maddenin öğrenmesi için, ***kendisini*** yeniden düzenleyerek arzu edilen fonksiyonu hesaplamada giderek daha da iyi olması lazım; yalnızca fizik kanunlarına uyarak.

Öğrenme sürecini aydınlatmak için, ilk olarak basit bir fiziksel sistemin nasıl π sayısının ve diğer sayıların basamaklarını öğrenebildiğine bakalım. Yukarıda pek çok vadisi olan bir yüzeyin (bkz. Görsel 2.3) nasıl bir bellek cihazı olarak kullanılabildiğini gördük: Mesela, vadinin en altındaki noktalardan biri $x = \pi =\approx 3.14159$ ise ve yakında başka vadiler yoksa, $x = 3$'e bir top koyarsanız sistemin topun en dibe kaymasına izin verip kalan ondalıkları hesaplayışını izlersiniz. Şimdi, bu yüzeyin yumuşak kilden yapıldığını ve beyaz bir sayfa olarak, tamamen düz başladığını düşünün. Eğer matematik sevdalıları devamlı olarak topu en sevdikleri sayılara koyarlarsa, yer çekimi yavaş yavaş bu konumlarda vadiler oluşturacaktır, bundan sonra da kil yüzey bu depolanmış hatıraları geri çağırmakta kullanılır. Diğer bir deyişle, kil yüzey π gibi sayıların basamaklarını hesaplamayı *öğrenmiştir.*

Beyinler gibi diğer fiziksel sistemler, aynı fikirden yola çıkarak çok daha etkili biçimde öğrenebilir. John Hopfield yukarıda bahsedilen bağlı nöronlar ağının benzer biçimde öğrenebildiğini göstermişti: Eğer devamlı olarak belli durumlara koyarsanız, yavaş yavaş bu durumları öğrenecek ve herhangi bir yakın durumdan onlara dönecektir. Eğer aile üyelerinizin her birini defalarca görürseniz, neye benzediklerine dair hafızaları onlarla ilgili herhangi bir şey sebebiyle tetiklenebilir.

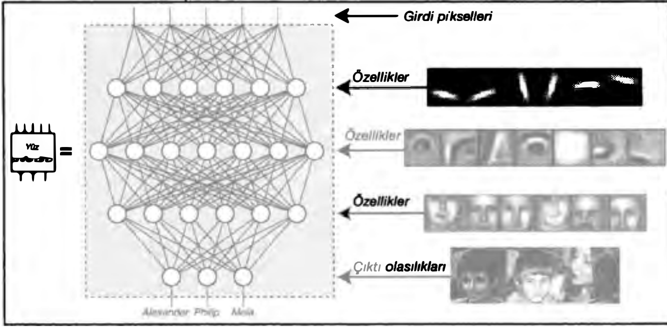
Sinirsel ağlar hem biyolojik hem de yapay zekâyı dönüştürdüler ve *makine öğrenimi* (deneyim yoluyla iyileşecek algoritma çalışması) olarak bilinen YZ alt alanına da hâkim olmaya başladılar. Böyle ağların nasıl öğrenebildiğini ayrıntılı biçimde incelemeden önce, nasıl hesapladıklarına bakalım. Bir sinirsel ağ birbirinin davranışını etkileyebilecek bağlantılı bir grup nörondur. Beyniniz galaksimizdeki yıldızlar kadar nörona sahiptir: yaklaşık olarak yüz milyar. Ortalama olarak, her bir nöron *sinaps* adı verilen bağlantı noktalarıyla yaklaşık bin tanesine bağlıdır ve beyninizdeki bilginin çoğunu kodlayan da bu yaklaşık yüz trilyon sinaps bağlantılarının kuvvetidir.

Bir sinirsel ağı şematik olarak, çizgiler sinapsları noktalar da onlarla bağlanmış nöronları temsil edecek şekilde çizebiliriz (bkz. Görsel 2.9). Gerçek dünya nöronları bu şematik illüstrasyona hiç benzemeyen oldukça komplike elektro-kimyasal cihazlardır: Akson ve dendrit gibi isimlere sahip farklı kısımları vardır, geniş bir çeşitlilikte çalışan pek çok farklı tür nöron bulunur ve bir nörondaki elektriksel aktivitenin diğer nöronları nasıl ve ne zaman etkilediği halen mevcut çalışmaların konusudur. Ancak YZ araştırmacıları sinir ağlarının, tüm bu karmaşıklıklar göz ardı edildiğinde bile pek çok karmaşık görevde insan seviyesi performansa ulaşabildiğini ve hepsi birbirinin aynı olan ve basit kurallara uyan basit simüle edilmiş olanlarla gerçek biyolojik nöronların yer değiştirebildiğini göstermiştir. Şu an böyle bir

yapay sinir ağı için en popüler model her nöronun durumunu ve her sinapsın kuvvetini tek bir rakamla temsil eder. Bu modelde, her nöron durumunu, düzenli zaman aralıklarıyla günceller. Bunu yaparken önce tüm bağlı nöronlardan gelen girdilerin ortalamasını alır, sinaptik kuvvetlerine göre onları tartar, opsiyonel olarak bir sabit ekler ve sonra da bir sonraki durumunu hesaplaması için sonuca *aktivasyon fonksiyonu* denen şeyi uygular.* Sinirsel bir ağı bir fonksiyon olarak kullanmanın en kolay yolu, Görsel 2.9'daki gibi, bilginin tek yöne aktığı bir biçimde *ileribildirim* yapmaktır. Yani fonksiyonu bir katman nörona yukarıdan bağlamak ve bir katman nöron olarak aşağıdan çıktıyı almak.

Bu basit yapay sinirsel ağların başarısı madde bağımsızlığının başka bir örneğidir: sinirsel ağların yapılarının düşük seviye temel detaylarından görünürde bağımsız yüksek hesaplama güçleri vardır. Gerçekten de George Cybenko, Kurt Hornik, Maxwell Stinchcombe ve Halbert White 1989'da muazzam bir şeyi kanıtladılar: Böyle basit sinirsel ağlar yalnızca bu sinaps kuvvet numaralarını uygun biçimde ayarlamak suretiyle rastgele biçimde *herhangi bir* fonksiyonu doğru hesaplayabildikleri için *evrenseldirler.* Diğer bir deyişle, evrim biyolojik nöronlarımızı gerekli olduğu için değil, daha verimli olduğu için böyle karmaşık yapmıştı. İnsan mühendislerin aksine evrim, anlaması kolay ve basit tasarımları ödüllendirmez.

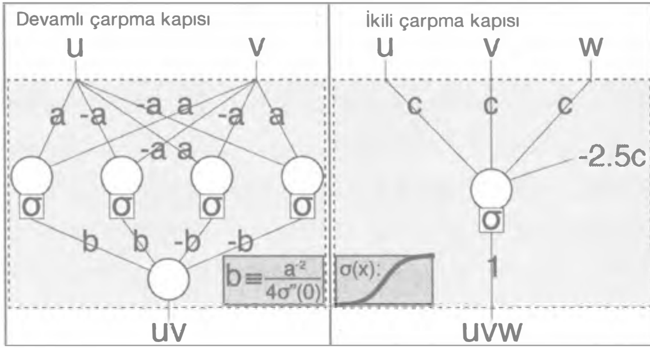
* Matematiği seviyorsanız, bu aktivasyon fonksiyonunun iki popüler seçimi sigmoid fonksiyonu $\sigma(x) = 1/(1 + e^{-x})$ ve ramp fonksiyonudur $\sigma(x) = \max\{0, x\}$. Gerçi, lineer (düz bir çizgi) olmadığı sürece neredeyse her fonksiyonun uygun olduğu kanıtlanmıştır. Hopfield'ın ünlü modeli $\sigma(x) = -1$ eğer $x < 0$ ise ve $\sigma(x) = 1$ eğer $x \geq 0$ ise denklemlerini kullanır. Eğer nöron durumları bir vektörde depolandıysa ağ sinaptik eşlenikleri içeren bir matrisle o vektörün çarpılması ve sonrasında da tüm elemanlara σ fonksiyonunun uygulanması ile güncellenir.



Görsel 2.9: Bir nöron ağı bir NAND kapıları ağı gibi fonksiyonları hesaplayabilir. Mesela, yapay sinirsel ağlar farklı görsel piksellerinin parlaklığını temsil eden rakamları girdi olarak alacak ve görüntünün farklı insanları tasvir etme olasılığını yansıtan rakamları çıktı olarak verecek şekilde eğitilmiştir. Burada, her yapay nöron (yuvarlak) yukarıdaki bağlantılar (çizgiler) yoluyla ona gönderilen rakamların toplamını hesaplar, basit bir fonksiyon uygular ve sonucu aşağıya, yüksek seviye özellikleri hesaplayan sonraki katmana yollar. Tipik yüz tanımlama ağları yüz binlerce nöron içerir; bu şekil, açıklık için çok azını göstermektedir.

Bunu ilk öğrendiğimde, böylesine basit bir şeyin böylesine rastgele karmaşık bir şeyi nasıl hesaplayabildiğine şaşırmıştım. Mesela, tek yapabildiğiniz şey ağırlık toplamı hesaplamak ve tek bir sabit fonksiyon uygulamakken, çarpma gibi basit bir işlemi bile nasıl yapabilirsiniz ki? Bunun nasıl çalıştığına dair biraz daha bilgi almak isterseniz, Görsel 2.10 beş nöronun iki rastgele rakamı ve tek bir nöronun üç biti nasıl çarptığını gösteriyor.

Teoride rastgele büyük bir sinirsel ağla her şeyi hesaplayabileceğinizi kanıtlayabilseniz bile, kanıt bunu mantıklı büyüklükteki bir ağla *pratikte* yapıp yapamayacağınıza dair bir şey söylemez. Aslında bunun üzerine ne kadar çok düşünürsem, sinirsel ağların çok iyi çalışması konusunda o kadar şaşkınlığa düşerim.



Görsel 2.10: Madde nasıl çarpma yapar. Görsel 2.7'deki gibi NAND kapılarını kullanarak değil nöronlarla. Detayları takip etmemize gerek yoktur. Ana nokta, nöronların (yapay ya da biyolojik) matematik işlemi yapabilmesinden öte çarpmanın NAND kapılarından daha az nörona ihtiyaç duymasıdır. *İleri derece matematik hayranları için ek detaylar:* Daireler toplama yapar, kareler σ fonksiyonunu uygular ve çizgiler onları işaretleyen sabitlerle çarpar. Girdiler (solda) gerçek rakamlar ve (sağda) bitlerdir. Çarpma $a \rightarrow 0$ (sol) ve $c \rightarrow 5$ (sağ) iken rastgele biçimde doğru olur. Sol ağ orijinde kavisli olan (yani ikinci türevi σ "(0) ≠ 0) herhangi bir σ(x) fonksiyonu için geçerlidir, bu da Taylor açılımı σ(x) ile kanıtlanabilir. Sağdaki ağ, fonksiyon σ(x)'in x çok küçük ve çok büyük olduğunda 0 ve 1'e yaklaşmasını gerektirir, bu da yalnızca $u+v+w=3$ iken $uvw=1$ olduğunda görülür (Bu örnekler öğrencilerim Henry Lin ve David Rolnik'le yazdığım bir makaleden: "Why Does Deep and Cheap Learning Work So Well?", şuradan erişebilirsiniz: http://arxiv.org/abs/1608.08225).

Örnek olarak, megapiksellik griölçek görüntülerini iki kategoriye, mesela kedi veya köpeklere ayırmak istediğimizi düşünelim. Eğer her milyon piksel, diyelim 256 değer alıyorsa, $256^{1000000}$ mümkün görüntü vardır ve her biri için, bir kediyi tasvir etme olasılığını hesaplamak istiyoruz. Bu da bir resim girdisi alan ve olasılık çıktısı veren bir rastgele fonksiyonun $256^{1000000}$ olasılıklı bir listeyle, yani evrenimizdeki atomlardan (yaklaşık 10^{78}) daha fazla olasılıkla tanımlandığı anlamına gelir. Yine de binler ya da milyonlarca parametrelere sahip sinirsel ağlar bir şekilde böyle

sınıflandırma görevlerini yerini getirmeyi başarır. Başarılı sinirsel ağlar az parametreye ihtiyaç duyma anlamında nasıl "ucuz" olabilir? Sonuçta, evrenimize sığacak kadar küçük bir sinirsel ağın, atayabileceğiniz tüm hesaplama görevlerinin komik derecede küçük bir kısmında başarıya ulaşarak neredeyse tüm fonksiyonları tahmin etmekte başarısız olacağını kanıtlayabiliriz.

Öğrencim Henry Lin'le bu ve benzeri gizemler üzerine pek fazla eğlenceli kafa karışıklığı yaşadık. Hayatımda minnettar olduğum şeylerden biri de muhteşem öğrencilerle çalışma imkânına sahip olmaktır ve Henry de onlardan biridir. Onunla çalışmak isteyip istemeyeceğimi sormak için ofisime ilk geldiğinde, kendi kendime bu soruyu benim sormamın daha doğru olacağını düşünmüştüm: Shreveport, Louisianalı alçak gönüllü, dost canlısı ve parlak gözlü bu çocuk, sekiz bilimsel makale yazmış, Forbes 30-Otuz Yaş Altı ödülünü kazanmış ve bir milyonun üzerinde kişinin izlediği bir TED konuşması yapmıştı ve yalnızca yirmi yaşındaydı! Bir yıl sonra, şaşırtıcı bir sonucu olan bir makale yazdık beraber: Sinirsel ağların neden çok iyi çalıştığı sorusu yalnızca matematikle cevaplanamazdı çünkü cevabın bir kısmı fizikte yatıyordu. Fizik kurallarının bize dayattığı ve bizi hesaplama konusunda meraklandıran fonksiyon sınıfının oldukça küçük bir sınıf olduğunu çünkü hâlâ daha tam olarak anlayamadığımız sebeplerden ötürü fizik kurallarının oldukça basit olduğunu bulduk. Dahası, sinirsel ağların hesaplayabildiği fonksiyonların o küçük kısmı fiziğin bizi meraklandıran küçük kısmına çok benziyordu! Daha önce yaptığımız ve derin öğrenme sinirsel ağlarının (eğer pek çok katmana sahiplerse "derin" denir) bu ilgimizi çeken fonksiyonların büyük kısmı için sığ olanlara göre çok daha verimli olduğunu gösteren çalışmamızı da genişlettik. Mesela, başka bir muhteşem MIT öğrencisi David Rolnick'le *n* sayısının çarpılması gibi basit bir görev tek bir katmanlı bir ağ için $2n$ sayıda nörona

ihtiyaç duyarken derin bir ağda $4n$ nörona ihtiyaç duyar. Bu da YZ araştırmacıları arasında sinirsel ağların neden büyük bir merak unsuru olduğunu, ayrıca beyinlerimizde neden sinirsel ağlar geliştirdiğimizi de açıklar: Beyinlerimizi geleceği tahmin edecek şekilde geliştirseydik, o zaman fiziksel dünyada önemli olacak bu hesaplama problemlerinde başarılı bir hesaplama mimarisi geliştirmemiz mantıklı olurdu.

Sinirsel ağların nasıl çalışıp hesap yaptığını keşfettiğimize göre, nasıl öğrendikleri sorusuna geri dönelim. Bilhassa sinirsel bir ağın sinapslarını güncelleyerek hesaplamada iyileşmesi nasıl gerçekleşir?

1949 tarihli ufuk açıcı kitabı *The Organization of Behavior: A Neuro-psychological Theory*'de, Kanadalı psikolog Donald Hebb yan yana iki nöron sık sık aynı anda aktifse ("ateşleniyorsa"), sinaptik eşleniklerin birbirlerini tetiklemeyi öğrenecek kadar güçleneceğini ileri sürer. Bu fikir popüler "birlikte ateşlenen, birlikte bağlanır" sloganıyla dile getirilir. Gerçek beyinlerin nasıl öğrendiğinin detayları hâlâ anlaşılmaktan uzak olsa ve araştırmalar cevapların çoğu durumda çok daha karmaşık olduğunu gösterse de bu basit öğrenme kuralı (Hebbian öğrenme) bile sinirsel ağların ilginç şeyler öğrenmeye izin verdiğini göstermektedir. John Hopfield, Hebbian öğrenmenin aşırı basit yapay sinirsel ağın devamlı olarak onlara maruz kalarak karmaşık birçok hafızayı depolayabilmesini sağladığını göstermişti. Bilgiye maruz kalarak öğrenmeye yapay sinirsel ağlarda (ya da yetilerin öğretildiği hayvanlar ve insanlarda) idman denir fakat "çalışma", "eğitim" ya da "deneyim" de denebilir. Bugünün YZ sistemlerine güç sağlayan yapay sinirsel ağlar Hebbian öğrenmeyi "geriyayılım" ve "stokastik gradyan iniş" gibi isimlere sahip daha karmaşık öğrenme kuralları ile değiştirme eğilimindedir ancak temel fikir yine aynıdır: Sinapsların zaman içinde güncellendiği bir fizik kuralına benzer, basit

bir deterministik yasa vardır. Sanki sihir gibi, bu basit kural eğer idman yüksek oranda veriyle gerçekleştirilirse inanılmaz derecede karmaşık hesaplamaları sinirsel ağın öğrenmesini mümkün kılar. Tam olarak beynimizin hangi öğrenme kurallarını kullandığını bilmiyoruz ancak cevap ne olursa olsun, fizik kurallarını ihlal ettiklerine dair elimizde bir kanıt yok.

Çoğu dijital bilgisayarın işlerini çoklu adımlara bölerek ve hesaplama modüllerini pek çok kez kullanarak verimlilik kazandırması gibi, pek çok yapay ve biyolojik sinirsel ağ da aynı şeyi yapar. Beyinler de ileribildirim sinir ağları yerine bilgisayar bilimcilerin *yineleyen* dediği, bilginin tek bir yön yerine çoklu yöne ilerlediği çıktı akışının da bir sonraki adımın girdisi olabildiği kısımlara sahiptir. Bir dizüstü bilgisayarın mikroişlemcisindeki mantık kapıları ağı da bu anlamda yineleyendir. Geçmiş bilgisini devamlı olarak yeniden kullanmayı sürdürür ve klavye, *trackpad*, kamera *vb.*'den gelen yeni bilginin devam eden hesaplamayı etkilemesine izin verir. Bunun karşılığında da bir ekrana, hoparlöre, yazıcı ya da kablosuz ağa gidecek bilgi çıkışını belirler. Benzer biçimde, beyninizdeki nöron ağı da yineleyendir. Gözleriniz, kulaklarınız ve diğer duyularınızdan gelen bilgi girdisinin devam eden hesaplamayı etkilemesine izin vererek kaslarınıza gidecek bilgi çıkışını belirler.

Öğrenmenin tarihi yaşamın tarihi kadar eskidir çünkü her kendini kopyalayan organizma ilginç kopyalama ve bilgi işleme eylemleri gerçekleştirir, bu davranışlar bir şekilde öğrenilmiştir. Yaşam 1.0 döneminde ise, organizmalar yaşam süreleri dâhilinde öğrenmediler. Bilgi işleme ve tepki gösterme kuralları miras aldıkları DNA'ları tarafından belirlenirdi, bu yüzden de tür seviyesinde yavaşça gerçekleşen öğrenme yalnızca Darwinci evrimle nesiller boyunca gerçekleşiyordu.

Yaklaşık yarım milyar yıl önce, dünyadaki bazı gen hatları sinirsel ağlara sahip hayvanları yaşamları süresince deneyimlerden davranış öğrenebilecek hale sokmanın bir yolunu keşfetti. Yaşam 2.0 geldi ve dramatik olarak hızlı öğrenme ve rekabeti zekâsıyla alt etme yetisi sayesinde, tüm dünyada kontrol edilemeyen bir yangın gibi yayıldı. Bölüm 1'de de gördüğümüz gibi, yaşam artan bir hızla öğrenerek daha da iyi hale geldi. Belli bir maymun benzeri tür bilgi edinmekte öylesine hünerli bir beyin geliştirdi ki aletleri nasıl kullanacağını, ateşi nasıl yakacağını, nasıl konuşacağını ve karmaşık bir küresel toplumu nasıl kurabileceğini öğrendi. Bu toplumun kendisi her yenilik bir sonrakini müjdelediği için giderek artan bir hızla hatırlayan, hesaplayan ve öğrenen bir sistem olarak görülebilir: yazı, matbaa, modern bilim, bilgisayarlar, internet ve dahası. Geleceğin tarihçileri bu yenisini önceleyen icatlar listesinin son halkasına kim bilir ne ekleyecek? Benim tahminim bunun yapay zekâ olacağı yönünde.

Bildiğimiz gibi, bilgisayar belleği ve hesaplama gücündeki devasa gelişmeler (Görsel 2.4 ve Görsel 2.8) yapay zekâda olağanüstü ilerlemeyi sağladı ancak makine *öğreniminin* olgunlaşması epey uzun sürdü. IBM'in Deep Blue bilgisayarı 1997'de satranç şampiyonu Garry Kasparov'u yendiğinde, ana avantajı bellek ve hesaplamadaydı, öğrenmede değil. Hesaplama zekâsı da bir grup insan tarafından yapılmıştı ve Deep Blue'nun yaratıcılarını yenebilmesinin temel sebebi hızlı hesap yapabilmesi ve böylelikle daha fazla potansiyel pozisyonu analiz edebilmesiydi. IBM'in Watson bilgisayarı *Riziko* isimli yarışma programında dünya şampiyonu insanı yendiğinde, yine üstün hafıza ve hızla özel olarak programlanmış yetilere dayanıyordu, öğrenmeye değil. Aynı şey ayaklı devinimden sürücüsüz araçlar ve kendi kendine inebilen roketlere kadar robotikteki pek çok atılım için geçerlidir.

Aksine, en son gerçekleşen YZ atılımlarının pek çoğunun ardında makine *öğrenimi* yatmaktadır. Görsel 2.11'i ele alalım örneğin. Neyin fotoğrafı olduğunu söylemek sizin için kolaydır ancak bir görüntünün tüm piksellerinin renklerini girdi olarak alan ve "Bir grup genç insan frizbi oynuyor" yazısını çıktı olarak veren bir fonksiyon programlamak on yıllardır dünyanın tüm YZ araştırmacılarının başaramadığı bir şeydi. Ilya Sutskever tarafından yönetilen bir Google ekibi tam olarak bunu 2014'te yapmayı başardı. Farklı bir set piksel rengi girdisine, "Bir grup fil kuru otlardan oluşan bir alanda yürüyor," diye yine doğru cevabı verebiliyordu. Bunu nasıl yapmışlardı? Deep Blue'da olduğu gibi, frizbiyi, yüzleri ve benzeri şeyleri tespit edecek elle girilmiş algoritmalarla mı? Hayır, fiziksel dünya ve içeriği hakkında hiçbir bilgiye sahip olmayan görece basit bir sinirsel ağ yaratıp bu ağı büyük oranlarda veriye maruz bırakarak öğrenmesini sağlayarak. YZ vizyoneri Jeff Hawkins 2004'te, "Hiçbir bilgisayar... Bir fare kadar iyi göremez," diye yazmıştı ama bu günler geride kalmıştı.

Çocuklarımızın nasıl öğrendiğini tam olarak anlayamadığımız gibi böyle sinirsel ağların da nasıl öğrenebildiğini ve neden ara sıra başarısız olduklarını tam olarak anlayamıyoruz. Ancak açık olan şey zaten yüksek oranda faydalı olmaları ve derin öğrenmede bir dizi yatırımı tetikliyor olmalarıdır. Derin öğrenme bilgisayar görüşünü, el yazısı transkripsiyonundan sürücüsüz araçların gerçek zamanlı video analizine kadar farklı açılardan değiştirmiştir. Bilgisayarların konuşma dilini metne dökme ve başka dillere çevirme yetilerinde de devrim yaratmıştır. Hatta bunu gerçek zamanlı bile yapabilir. Bu yüzden de Siri, Google Now ve Cortana gibi kişisel dijital yardımcılarla konuşabiliyoruz. Bir web sitesine insan olduğumuzu kanıtlamak zorunda olduğumuz o can sıkıcı CAPTCHA bilmeceleri makine öğrenme teknolojisinin yapabileceklerinin önünde kalabilmek adına sürekli daha

da zorlaşıyor. 2015 yılında, Google DeepMind bir çocuk gibi düzinelerce bilgisayar oyununda ustalaşabilen –hem de hiçbir talimat olmadan– ve hatta kısa sürede herhangi bir insandan daha da iyi oynamayı öğrenen bir derin öğrenme kullanan bir YZ sistemi yayımladı. 2016 yılında, aynı şirket, farklı tahta pozisyonlarının kuvvetini ölçebilen bir derin öğrenme kullanan, Go oynayabilen bir bilgisayar sistemi olan AlphaGo'yu inşa etti ve bu sistem dünyanın en güçlü Go şampiyonunu yendi. Bu ilerleme verimli bir döngü yaratarak daha fazla kaynak ve yeteneği YZ araştırmasına çekti, bu da daha fazla ilerlemeye yol açtı.

Bu bölümü zekânın doğası ve şu ana kadar gelişimine ayırdık. Makinelerin bizi *tüm* bilişsel görevlerde geçmesi ne kadar zaman alacak? Net olarak bilmiyoruz ve cevabın "asla" olabileceği ihtimaline karşı da açık olmalıyız. Ancak bu bölümün temel mesajı, *gerçekleşeceği* ve hatta bunun bizim yaşam döngümüz içinde olabileceği ihtimalini de değerlendirmek zorunda olduğumuzdur. Sonuçta, madde fizik kurallarına uygun biçimde düzenlendiğinde hatırlayabiliyor, hesaplayabiliyor ve öğrenebiliyor – hem de bu maddenin biyolojik olması da gerekmiyor. YZ araştırmacıları büyük şeyler söz vermek ve bunları yerine getirememekle suçlanır ama adil olmak gerekirse, bazı eleştirmenlerinin de çok iyi performansı olduğu söylenemez. Bazıları kale direklerini sürekli ileri taşıyarak zekâyı bilgisayarların halen daha yapamadığı şeyler ya da bizi etkileyen şeyler olarak tanımlıyor. Makineler şu an aritmetik, satranç, matematiksel teorem kanıtlama, hisse senedi seçme, görsel yazılama, araba sürme ve oyun oynama, Go, konuşma sentezi, konuşma transkripsiyonu, çeviri ve kanser teşhisi konularında iyi ya da mükemmel durumda ancak bazı eleştirmenler "Elbette ama bu *gerçek* zekâ değil!" diyebiliyorlar. Hatta gerçek zekânın yalnızca Moravec'in manzarasında (Görsel 2.2) henüz su altında kalmamış dağ tepeleri olduğunu iddia edebilirler. Tıpkı geçmişte

bazılarının –su seviyesi devamlı yükselirken– görüntü tanıma ve Go'nun sayılması gerektiğini ileri sürerken yaptıkları gibi.

Suyun en azından bir süre daha yükseleceğini varsayarsak, YZ'nin toplum üzerindeki etkisinin de artacağını söyleyebiliriz. YZ tüm görevler için insan seviyesine gelmeden çok önce, bize program hataları, hukuk, silahlar ve iş gücü gibi konularda baş döndürücü fırsatlar ve zorluklar yaşatacaktır. Bunlar nelerdir ve onlara en iyi nasıl hazırlanabiliriz? Bunu da bir sonraki bölümde inceleyelim.



Görsel 2.11: "Bir grup genç insan frizbi oynuyor" – bu alt başlık insan, oyun ya da frizbiye dair hiçbir kavrayışı olmayan bir bilgisayar tarafından yazıldı.

SONUÇ:

- Karmaşık hedeflere ulaşmak olarak tanımlanmış zekâ tek bir IQ değeriyle değil, tüm hedefler arasında bir yeti spektrumuyla ölçülebilir.
- Bugünün yapay zekâsı her sistem yalnızca oldukça spesifik hedefleri yerine getirdiği için, *dar* olma eğilimindedir. İnsan zekâsı ise olağanüstü bir biçimde *geniştir.*

- Bellek, hesaplama, öğrenme ve zekâ; soyut, fiziksel varlığı olmayan ve uhrevi gibi gelir çünkü hepsi de *maddeden bağımsızdır:* altyapısında var olan maddenin detaylarını yansıtmayan ya da onlara bağlı olmayan bir şekilde, hayatı kendi başına sürdürebilir.
- Herhangi bir madde bütünü, pek çok farklı denge durumuna sahip olduğu sürece *hafıza* için öz olabilir.
- Herhangi bir madde, herhangi bir fonksiyonu uygulamak için bir araya getirilebilecek belirli evrensel yapı taşlarına sahip olduğu sürece *hesaplama* özü, *computronium* olabilir. NAND kapıları ve nöronlar böyle evrensel "hesaplama atomları" için iki önemli örnektir.
- Sinirsel ağ, *öğrenme* için güçlü bir öz maddedir çünkü yalnızca fizik kurallarına uyarak, kendini arzu edilen hesaplamaları gerçekleştirmekte daha iyi hale getirecek şekilde yeniden düzenleyebilir.
- Fizik kurallarının çarpıcı basitliğinden ötürü, biz insanlar tüm hayal edilebilir hesaplama problemlerinin yalnızca ufacık bir kısmıyla ilgileniyoruz ve sinirsel ağlar bu ufak kesimi çözmekte olağanüstü biçimde iyidir.
- Teknoloji iki katı güçlü hale geldiğinde, kendinden iki kat daha güçlü olacak teknolojiyi tasarlamak ve inşa etmekte kullanılabilir ve Moore yasası ruhu ile tekrar eden bir biçimde ikiye katlama kabiliyetini tetikleyebilir. Bilgi teknolojisinin maliyeti şu anda yüzyıl boyunca iki yılda bir yarıya inerek bilgi çağını oluşturmuştur.
- Eğer YZ ilerlemesi devam ederse, YZ tüm yetilerde insan seviyesine ulaşmadan çok önce, bize program hataları, hukuk, silahlar ve iş gücü gibi konularda baş döndürücü fırsatlar ve zorluklar yaşatacaktır. Bunları sonraki bölümde inceleyeceğiz.

Bölüm 3:



Yakın Gelecek: Atılımlar, Program Hataları, Hukuk, Silahlar ve İşler

Eğer yönümüzü kısa süre içinde değiştirmezsek, gittiğimiz yere varacağız.

Irwin Corey

Günümüzde ve bu çağda insan olmak ne demektir? Mesela, kendimizde değer verdiğimiz, bizi diğer yaşam formları ve makinelerden ayıran şey nedir? İnsanlar bazılarımıza iş verirken neye göre yapar bunu? Bu sorulara cevabımız ne olursa olsun, teknolojinin yükselişi gitgide bu cevapları değiştirmek zorundadır.

Mesela beni ele alalım. Bir bilim insanı olarak, yaratıcılığı ve önseziyi geniş bir kapsamdaki çözülmemiş sorunları halletmekte ve dili de keşfettiklerimi paylaşmakta kullanmak anlamında kendi hedeflerimi kurmuş olmaktan gurur duyarım. Neyse ki benim durumumda toplum bana bunu bir iş olarak yapmam için para ödemeye razıdır. Yüzlerce yıl önce, bunun yerine, tıpkı birçok başkası gibi, kimliğimi bir çiftçi ya da zanaatkâr olmak etrafında kurabilirdim ancak teknolojinin ilerlemesi bu meslekleri iş gücünün çok küçük bir kısmını oluşturmaya itti. Bu da bugün artık

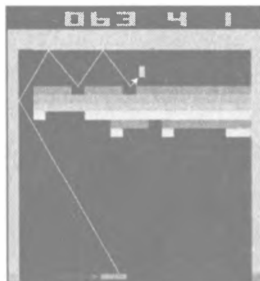
kimliğimizi çiftçilik ve zanaat etrafında inşa etmenin herkes için mümkün olmadığı anlamına geliyor.

Kişisel olarak, bugünün makinelerinin beni kazma ya da örme gibi el yetilerinde geçmesinden rahatsız olmuyorum çünkü bunlar ne benim hobilerim ne gelir kaynağım ne de kendime verdiğim değeri belirleyen şeyler. Gerçekten de bu yetilerim hakkında sahip olacağım herhangi bir hayal, sekiz yaşımda neredeyse kaldığım bir örme dersi almak zorunda olduğumda ve bana acıyan beşinci sınıftan bir arkadaşımın yardımıyla ancak projeyi tamamlayabildiğimde suya düşmüştü. Fakat teknoloji iyileşmeyi sürdürdükçe, YZ'nin yükselişi eninde sonunda bugün kendime verdiğim değeri belirleyen ya da iş piyasasında değeri olan yetilerimi de gölgede bırakırsa? Stuart Russell, bazı YZ araştırmacısı arkadaşlarıyla birlikte YZ'nin yıllar boyunca yaptığını göremeyeceklerini düşündükleri bir şeyi yapmasına tanıklık ettiklerinden bir, "Hadi be!" anı yaşadıklarını söylemişti. Aynı şekilde, size kendi, "Hadi be!" anlarımdan bazılarını ve onları insan yetilerinin yakın zaman içinde gölgede kalacağına dair işaretler olarak nasıl değerlendirdiğimi anlatayım.

Atılımlar

Derin Takviyeli Öğrenme Üstlenicileri

2014 yılında bir DeepMind YZ sisteminin bilgisayar oyunu oynamayı öğrenmesine dair bir video izlerken ağzım açık kalmıştı. YZ, gençlik yıllarımdan sevgiyle andığım klasik bir Atari oyunu olan Breakout'u (bkz. Görsel 3.1) oynuyordu. Amaç bir paleti hareket ettirip topu bir tuğla duvara çarptırmaktı; ne zaman bir tuğlaya çarparsanız tuğla kaybolur ve puanınız artar.



Görsel 3.1: Atari oyunu Breakout'u oynamayı sıfırdan öğrendikten ve derin takviyeli öğrenmeyi skorunu maksimize etmekte kullandıktan sonra, DeepMind YZ optimal stratejiyi keşfetti: tuğla duvarın en sol tarafına bir delik açmak ve topu onun arkasına atıp zıplatarak hızla puan toplamak. Top ve paletin önceki hareketlerini gösteren oklar çizdim.

Geçmişte kendi kendime bilgisayar oyunları yazmışlığım vardır ve Breakout'u oynayacak bir program yazmanın çok da zor olmadığını biliyordum. Ancak DeepMind takımının yaptığı şey bu değildi. Onun yerine, bu oyun hakkında hiçbir şey bilmeyen –ya da başka bir oyun hakkında hatta oyun, palet, tuğla ya da top gibi *kavramlar* hakkında bile hiçbir şey bilmeyen- boş sayfa bir YZ yaratmışlardı. YZ'nin bildiği tek şey düzenli aralıklarla ona beslenen uzun rakam listeleriydi: mevcut skor ve bizim (ama YZ'nin değil) ekranın farklı alanlarının nasıl renklendiğine dair tanımlamalar olarak göreceğimiz uzun bir sayılar listesi. YZ'ye sadece skoru, düzenli aralıklarla bizim (ancak YZ'nin değil) hangi tuşlara basılması gerektiğini gösteren kodlar olarak göreceğimiz sayıları çıktı olarak vererek maksimize edilmesi söylenmişti.

Başta YZ berbat oynuyordu: paleti sağa sola rastgele hareket ettiriyor, neredeyse her seferinde topu ıskalıyordu. Bir süre sonra, paleti topa doğru hareket ettirmenin iyi bir fikir olduğunu anlamaya başladı, yine de çoğu sefer topu ıskaladı. Ancak pratikle kendini geliştirdi ve kısa süre içinde oyunda benden daha iyi bir hale gelerek top ne kadar hızlı gelirse gelsin hata yapmadan onu

karşılayabildi. Sonrasında, ağzımı açık bırakan olay oldu işte: Duvarda bir delik açmak için sürekli sol üst köşeye hedefleyerek ve topu duvarın arkası ile ardındaki bariyer arasında sekmeye zorlamak suretiyle skoru maksimize etme stratejisini buldu. Bu çok zekice bir şey gibi gelmişti. Hakikaten, Demis Hassabis sonradan bana DeepMind'daki programcıların bu numarayı inşa ettikleri YZ'den öğrenene kadar bilmediklerini söylemişti. Size bunun videosunu verdiğim linkten izlemenizi öneririm.[1]

Burada bir şekilde rahatsız edici, insan benzeri bir özellik vardı: Hedefi olan bir YZ'nin o hedefe ulaşmakta giderek iyileşmeyi öğrenişini, zamanla da yaratıcılarını geçtiğini izliyordum. Bir önceki bölümde, zekâyı basitçe karmaşık hedeflere ulaşmak anlamında kullanmıştık. Bu anlamda DeepMind'ın YZ'si gözlerimin önünde daha da zeki oluyordu (gerçi yalnızca bu oyunu oynamak gibi dar bir anlamda). İlk bölümde, bilgisayar bilimcilerin *akıllı üstlenici* adını verdikleri şeyle karşılaşmıştır: Sensörlerden çevreleri hakkında bilgi toplayan, sonra da bu bilgiyi çevrelerine nasıl karşılık vereceklerine karar vermekte kullanan varlıklar. DeepMind'ın oyun oynayan YZ'si olabildiğince basit, tuğlalar, paletler ve toplardan oluşan sanal bir dünyada yaşıyor olsa da akıllı bir üstlenici olduğunu reddedemezdim.

DeepMind kısa süre içerisinde metotlarını yayımladı ve kodlarını paylaşarak *derin takviyeli öğrenme* adı verilen çok basit ama güçlü bir fikri kullandığını açıkladı.[2] Basit takviyeli öğrenme, davranışsal psikolojiden esinlenen, olumlu bir ödül almanın o şeyi yeniden yapma eğilimini artırdığı (veya tam tersi) bir klasik makine öğrenimi tekniğiydi. Bir köpeğin sahibinden teşvik ya da yiyecek alma ihtimali arttığında numaraları yapmayı öğrenmesi gibi DeepMind YZ'si paleti topu yakalayacak şekilde hareket ettirmeyi öğrenmişti çünkü bu daha fazla puan kazanma ihtimalini artırıyordu. DeepMind bu fikri derin öğrenmeyle birleştirdi:

Bir önceki bölümdeki gibi, derin sinirsel bir ağı klavyede izin verilen her bir tuşa basarak elde edilebilecek ortalama puanın ne olduğunu tahmin etme konusunda eğitti ve sonrasında da YZ oyunun mevcut anı içinde sinirsel ağın en çok ümit vadeden tuş olarak ileri sürdüğü tuşu seçti.

Bir insan olarak kendime değer verme fikrine en çok katkıda bulunan özelliklerimi sıraladığımda, geniş bir çapta çözülmemiş sorunu çözme yetisini ekliyorum. Tam tersine, Breakout oynamak ve başka hiçbir şey yapmamak oldukça dar bir zekâya tekabül eder. Bana göre DeepMind'ın atılımının asıl önemi, derin takviyeli öğrenmenin tamamen genel bir teknik olmasından gelir. Hakikaten, aynı YZ'ye kırk dokuz farklı Atari oyunu pratik ettirdiler ve Pong'dan Boks'a, Video Pinball'dan Space Invaders'a bu oyunların 29'unda insan deneklerden daha iyi oynamayı öğrenebildi.

Aynı YZ fikrinin kendini iki değil üç boyutlu olan daha modern oyunlarda da kanıtlamaya başlaması uzun sürmedi. Kısa sürede DeepMind'ın OpenAI'daki San Francisco merkezli rakipleri DeepMind'ın YZ'sinin ve diğer akıllı üstlenicilerin bir oyunmuş gibi tüm bir bilgisayarla etkileşime girmeyi pratik edebildiği Universe isimli bir platform yayınladılar: Her şeyi tıklayabiliyor, yazabiliyor, gezebildikleri yazılımları açıp çalıştırıyorlar, bazen web tarayıcısını açıp internette dolaşabiliyorlardı.

Derin takviyeli öğrenmenin geleceğine ve müteakip gelişmelere bakınca, görünürde bir son yoktur. Potansiyel sanal oyun dünyalarıyla sınırlı değil çünkü robotsanız, hayatın kendisi de bir oyun gibi görülebilir. Stuart Russell bana ilk HB anını Big Dog isimli robotun karla kaplı orman yamacından yukarı koşuşunu, yıllar boyunca çözmeye çalıştığı ayaklı devinim problemini incelikli bir biçimde çözüşünü izlerken yaşadığını söylemişti.[3] Bu kilometre taşına 2008'de ulaşılması, zeki programcıların ciddi çalışmalarıyla olmuştu. DeepMind'ın atılımından sonra, bir

robotun insan programcılardan yardım almadan kendi kendine yürümeyi öğrenmesi için bir tür derin takviyeli öğrenmeyi kullanmaması için hiçbir sebep kalmamıştır: İhtiyaç duyulan tek şey ilerleme katettiğinde puan veren bir sistemdir. Gerçek dünyadaki robotlar benzer bir şekilde yüzmeyi, uçmayı, pinpon oynamayı, kavga etmeyi ve neredeyse sonsuz sayıdaki motor görevleri insan programcıların yardımı olmadan öğrenme potansiyeline sahiptir. İşleri biraz hızlandırmak ve öğrenme sürecinde tıkanma ya da kendilerine zarar verme risklerini azaltmak için, öğrenmelerinin ilk birkaç aşamasını muhtemelen sanal gerçeklikte yapacaklardır.

Önsezi, Yaratıcılık ve Strateji

Benim için başka bir dönüm noktası DeepMind YZ sistemi AlphaGo'nun yirmi birinci yüzyılın başlarında dünyada en iyi oyuncu olarak kabul edilen Lee Sedol'e karşı beş oyunluk bir Go serisini kazanmasıydı.

İnsan Go oyuncularının bir noktada makineler tarafından tahtlarından edileceği bekleniyordu çünkü bu, satranç oyuncularının başına yirmi yıl kadar önce gelmişti. Ancak pek çok Go üstadı bunun bir on yıl daha süreceğini düşünüyordu. Bu yüzden de AlphaGo'nun zaferi onlar kadar benim için de bir dönüm noktasıdır. Nick Bostrom ve Ray Kurzweil YZ atılımlarının gelişini tahmin etmenin ne kadar zor olduğunun altını çiziyorlar. Zaten bu da Lee Sedol'ün ilk üç oyunu kaybetmesinden önce ve sonra verdiği röportajlardan da açıkça görünüyor:

- Ekim 2015: "Görünen seviyesini düşünürsek... Sanıyorum ki oyunu büyük farkla kazanacağım."

- Şubat 2016: "Google DeepMind YZ'sinin şaşırtıcı biçimde güçlü olduğunu ve giderek daha da güçlendiğini duydum ancak eminim ki en azından bu sefer kazanacağım."
- 9 Mart 2016: "Çok şaşkınım çünkü kaybedeceğimi düşünmemiştim."
- 10 Mart 2016: "Kendimi ifade etmekte zorlanıyorum... Şoktayım. İtiraf edeyim üçüncü oyun benim için pek de kolay olmayacak."
- 12 Mart 2016: "Çaresizmişim gibi hissettim."

Lee Sedol'le oynadıktan bir yıl sonra, daha da geliştirilmiş bir AlphaGo dünyanın en iyi yirmi oyuncusuyla oynadı ve tek bir maç bile kaybetmedi.

Bu şahsen benim için neden bu kadar önemliydi? Açıkçası yukarıda itiraf ettiğim gibi önsezi ve yaratıcılığı benim iki temel insani özelliğim olarak görüyorum ve şimdi açıklayacağım üzere, AlphaGo'nun ikisini de gösterdiğini hissediyorum.

Go oyuncuları 19'a 19 bir tahtada sırayla siyah ve beyaz taşları yerleştirirler (bkz. Görsel 3.2). Go oyununda evrenimizdeki atomlardan daha çok pozisyon mevcuttur, bu da sonraki tüm ilginç hamle dizilimlerini analiz etmeyi neredeyse imkânsız kılar. Oyuncular da bu yüzden bilinçli mantıksal çıkarımlarına eşlik etmek için bilinçaltı önseziye güvenirler. Uzmanlar neredeyse tekinsiz bir his geliştirerek hangi pozisyonun zayıf hangisinin güçlü olduğunu anlayabilirler. Son bölümde gördüğümüz üzere, derin öğrenmenin sonuçları bazen önseziyi çağrıştırır: Derin bir sinirsel ağ nasıl olduğunu açıklayamadan bir kedinin görselini resmedebilir. DeepMind takımı da bu yüzden derin öğrenmenin yalnızca kedileri değil, güçlü Go pozisyonlarını da tanıyabileceği fikri üzerine oynamışlardı. AlphaGo'ya işledikleri temel fikir

derin öğrenmenin önsezi gücüyle GOFAI'nin –derin öğrenme devrimi öncesinden kalma esprili bir deyim olan "Good Old-Fashioned AI", yani "O Eski İyi Yapay Zekâ" kısaltması– mantıksal gücünü bir araya getirmekti. Hem insan oyunlarından hem de AlphaGo'nun kendi klonuyla oynadığı oyunlardan devasa bir veri tabanını kullandılar ve bir derin sinirsel ağı her konumda beyazın en nihayetinde kazanacağı olasılığı tahmin etmek için eğittiler. Başka bir ağı da olası sonraki hamleleri tahmin etmek için eğittiler. Sonra bu iki ağı, mevcut en güçlü konuma yol açacak sıradaki hamleyi tanımlamak için potansiyel gelecek hareket dizilimleri listesini zekice tarayan bir GOFAI yöntemiyle birleştirdiler.

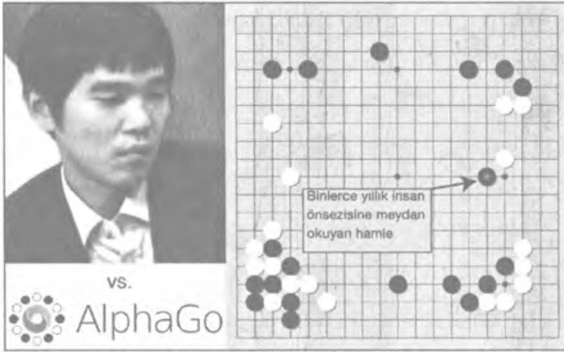
Zekâ ve önsezinin birleşimi, yalnızca güçlü değil, bazı durumlarda da yüksek ölçüde yaratıcı hamleleri doğurdu. Mesela, Go bilgeliğinin binlerce yılı oyunun başlarında bir taraftan üçüncü ya da dördüncü sırada bir hamle yapmanın en iyisi olduğunu söyler. İkisi arasında bir değiş tokuş vardır: Üçüncü sıraya yapılan hamle kısa sürede tahtanın kenarına doğru bölge kazanımı sağlarken dördüncü sıra uzun vadede merkeze doğru stratejik etkiyi sağlar.

İkinci oyunun otuz yedinci hamlesinde AlphaGo bu eski bilgeliğe karşı koyarak ve beşinci sıraya oynayarak Go dünyasını şoke etmişti (Görsel 3.2). Sanki uzun dönem planlama yetilerinde bir insandan daha iyiydi ve bu yüzden stratejik avantajı kısa dönemli kazancın üzerinde tutmuştu. Yorumcular şoke olmuştu ve Lee Sedol ayağa kalkmış ve geçici olarak odayı terk etmişti.[4] Hakikaten, yaklaşık elli hamle sonra, aşağı sol köşede yapılan mücadele taşmış ve otuz yedinci hamledeki siyah taşla birleşmişti! Bu hamle oyunu kazandıran şey olmuş, AlphaGo'nun beşinci sıra hamlesini Go tarihindeki en yaratıcı hamlelerden biri olarak efsaneleştirmişti.

Önseziye dayanan ve yaratıcı veçhelerinden ötürü Go, bir oyundan çok bir sanat formu olarak görülür. Antik Çin'de resim,

kaligrafi ve *qin* müziğiyle beraber dört "temel sanattan" biri olarak görülürdü ve Asya'da halen aşırı derecede popülerdi. Zaten AlphaGo ve Lee Sedol arasındaki ilk maçı üç yüz milyon kişi izlemişti. Sonuç olarak, Go dünyası neticeden çok etkilenmişti ve AlphaGo'nun galibiyetini insanlık için derin bir kilometre taşı olarak görmüştü. O zaman dünyanın en üst puanlı Go oyuncusu olan Ke Jie şunu demişti: "İnsanlık binlerce yıldır Go oynadı ama yine de, YZ'nin bize gösterdiği gibi daha yolun başındayız... İnsan ve bilgisayar oyuncuların birlikteliği yeni bir çağı açacak... Birlikte, insan ve YZ, Go'nun hakikatini bulacaktır." Böylesine verimli insan-makine ortaklığı, bilim gibi, YZ'nin insanlara nihai potansiyelimizi gerçekleştirmek ve kavrayışımızı derinleştirmekte yardımcı olacağı pek çok alanda ümit vadeder elbette.

Bana göre, AlphaGo ayrıca bize yakın geleceğe dair önemli bir ders de vermiştir: Derin öğrenme önsezisini GOFAI mantığıyla birleştirmek mükemmel bir *strateji* üretebilir. Go nihai strateji oyunlarından biri olduğu için YZ, oyun tahtalarının ötesinde en iyi insan stratejistlere meydan okumaya ya da onlara yardım etmeye hazırdır: yatırım stratejisi, politik strateji ve askerî strateji gibi. Böylesi gerçek dünya strateji sorunları tipik olarak insan psikolojisi, eksik bilgi ve rastgele olarak modellenmesi gereken faktörler sebebiyle karmaşık hale gelmiştir ancak poker oynayan YZ sistemleri çoktan bu zorlukların aşılamaz olmadığını göstermiştir.



Görsel 3.2: DeepMind'ın AlphaGo YZ'si sağdan beşinci sütuna, binlerce yıllık insan bilgeliğine meydan okuyan ve elli hamle kadar sonra Go efsanesi Lee Sedol'ün yenilgisinde hayati önem taşıyacak oldukça yaratıcı bir hamle yapmıştı.

Doğal Dil

YZ gelişmesinin beni yakın zamanda şoke ettiği başka bir alan da dildir. Gençliğimde seyahate düşkündüm ve başka kültür ve dillere olan merakım kimliğimin önemli bir parçasını oluşturdu. Hem İsveççe hem de İngilizce konuşacak biçimde yetiştirildim ve okulda Almanca ile İspanyolca öğrendim, iki evliliğimde Portekizce ve Rumence öğrendim ve sırf eğlencesine kendi kendime Rusça, Fransızca ve Mandarin öğrendim.

Fakat YZ bana yetişiyordu ve 2016'daki önemli bir keşiften sonra, Google'ın beyninin teçhizatı tarafından geliştirilen YZ sisteminden daha iyi çeviri yapabileceğim neredeyse hiçbir tembel dil kalmamıştı.

Kendimi ifade edebildim mi? Aslında şunu demeye çalışıyordum:

Fakat YZ bana yetişiyordu ve 2016'daki önemli atılımdan sonra, Google beyin takımının geliştirdiği YZ sisteminden daha iyi çevirisini yapabileceğim hiçbir dil kalmamıştı.

Fakat bunu önce İspanyolcaya çevirdim, sonrasında da birkaç yıl önce dizüstü bilgisayarıma yüklediğim bir uygulamayı kullanarak tekrar İngilizceye çevirdim. 2016'da, Google beyin takımı ücretsiz Google Translate hizmetlerini, derin yineleyen sinirsel ağları kullanacak şekilde güncellemişti ve eski GOFAI sistemleri üzerindeki iyileştirme dramatik olmuştu:[5]

YZ bana yetişiyordu ve 2016'daki atılımdan sonra, Google beyin takımının geliştirdiği YZ sisteminden daha iyi çevirisini yapabilecek hiçbir dil kalmamıştı.

Gördüğünüz gibi, "ben" zamiri İspanyolcaya çevrilip geri çevrilince kaybolmuştu ve maalesef bu da anlamın değişmesine yol açmıştı. Neredeyse ama tam değil! Fakat Google'ın YZ'sini savunmak adına, çözümlemesi zor, gereksiz uzun cümleler yazmakla eleştirilirim ve bu örnek için en kafa karıştırıcı ve karmaşık cümlelerimden birini seçtim. Daha tipik cümleler için, YZ'leri genelde kusursuz biçimde çeviri yapabiliyor. Sonuç olarak, ortaya çıktığında büyük bir heyecan yaratmıştı ve günde yüz milyonlarca insan tarafından kullanılacak kadar faydalı olmuştu. Dahası, konuşmadan metne ve metinden konuşmaya dönüşüm için derin öğrenmedeki geçmiş ilerlemelerin sayesinde, bu kullanıcılar şimdi akıllı telefonlarına bir dilde konuşup ardından çevrilen sonucu dinleyebiliyorlar.

Doğal dil işleme YZ'nin en hızlı ilerleyen alanlarından biri ve bence bu başarı daha büyük bir etki yaratacak çünkü dil, insan olmanın merkezindedir. YZ dilbilimsel tahminde ne kadar iyileşirse, mantıklı e-posta cevapları yazmakta ya da konuşmalara devam etmekte de o kadar iyileşir. Bu da dışarıdan bir gözlemcide insan düşüncesinin mevcut olduğu izlenimi yaratabilir. Derin öğrenme sistemleri de böylece bir makinenin bir insanı kendisinin insan olduğuna inandırabilecek kadar iyi biçimde yazılı bir

konuşmayı sürdürebileceği ünlü Turing testini geçmeye doğru bebek adımları atmaya başlar.

Ancak dil üzerine çalışan YZ'nin hâlâ çok yol alması gerekir. Bir YZ benden daha iyi çeviri yaptığında biraz üzüldüğümü itiraf etsem de şu ana kadar benim söylediğim hiçbir şeyi anlamlı biçimde *anlayamadığını* kendime hatırlattığımda tekrar iyi hissetmeye başlarım. Devasa veri setleri üzerinde çalışırken, YZ gerçek dünyada bu kelimelerin neye tekabül ettiğini bilmeden mevcut kelime ilişkilerini ve örüntüleri keşfeder. Mesela, her kelimeyi bazı belli başlı kelimelere ne kadar benzer olduğunu gösteren binlerce numaradan oluşan bir listeyle temsil edebilir. Sonra da buradan "kral" ve "kraliçe" arasındaki farkın "koca" ve "karı" arasındaki farka benzediğini çıkarsayabilir. Ancak yine de erkek ya da kadın olmanın ne demek olduğunu hatta dışarıda uzay, zaman ve maddeden oluşan bir fiziksel gerçeklik bulunduğunu asla bilemez.

Turing testi temel olarak aldatmayla ilgili olduğu için, gerçek yapay zekâdan çok insanların enayiliğini ölçmekle eleştirilmiştir. Aksine *Winograd Schema Challenge* ismi verilen rakip bir test doğrudan mevcut derin öğrenme sistemlerinin eksik bırakma eğiliminde olduğu genelgeçer kavrayışı hedef alır. Biz insanlar rutin olarak bir cümleyi çözümlerken, bir zamirin neye tekabül ettiğini bulmaya çalışırken gerçek dünya bilgisini kullanırız. Örneğin, tipik bir Winograd testi burada "onların" neye tekabül ettiğini sorar:

- "Belediye meclisi eylemcilere izin vermeyi reddetti çünkü onlar şiddetten korkuyordu."
- "Belediye meclisi eylemcilere izin vermeyi reddetti çünkü onlar şiddeti destekliyordu."

Bu tip sorulara cevap verilen geleneksel bir YZ yarışması vardır ve YZ'ler hâlâ çok kötü performanslar gösterir.[6] Tam olarak buradaki zorluk, neyin neye tekabül ettiğinin bulunmasıdır. Yukarıdaki örnekte İspanyolca yerine Çinceyi kullandığımda Google Translate'i bile mahveder:

Fakat YZ bana yetişti, 2016'da büyük bir kırılmadan sonra, neredeyse hiçbir dille, YZ sistemini Google beyin takımınca geliştirilmişten daha çevirebilirim.

Lütfen https://translate.google.com sitesinden kendiniz de deneyin ve Google'ın YZ'si gelişmiş mi görün! Gelişmiş olma ihtimali çok mümkün çünkü derin yineleyen sinirsel ağları GOFAI ile birleştirerek bir dünya modelini de içeren dil işleyen bir YZ kurmak için çok fazla ümit vadeden yaklaşım vardır.

Fırsatlar ve Zorluklar

Bu üç örnek yalnızca bir örneklemdir çünkü YZ pek çok önemli cephede hızla ilerlemektedir. Dahası, bu örneklerde yalnızca iki şirketten bahsetmeme rağmen, üniversitelerde ve diğer şirketlerdeki rakip araştırma grupları genelde pek de geride kalmamıştı. Apple, Baidu, DeepMind, Facebook, Google, Microsoft ve diğerleri öğrencileri, doktora sonrası öğrencileri ve hocaları çekmek için çok kazançlı teklifler sunarken, dünyanın dört bir yanındaki bilgisayar bilimi bölümlerinde gürültülü bir emme sesi duyulabilir.

Verdiğim örneklerden YZ tarihinin arada sırada olan atılımlarla kesilen duraklama dönemlerinden oluştuğu yanılgısına düşmemek önemlidir. Uzun bir süre boyunca oldukça istikrarlı bir gelişim olduğunu görüyorum. Medya bunu faydalı bir ürün ya da hayal gücünü tetikleyen bir uygulama eşiği aşıldığında atılım olarak lanse eder. Bu yüzden, ben de canlı YZ ilerlemesinin pek

çok yıl boyunca devam edeceğine inanıyorum. Dahası, bir önceki bölümde gördüğümüz üzere bu sürecin YZ pek çok görevde insan yetilerine ulaşana kadar devam etmemesi için hiçbir sebep yoktur.

Bunun sonucunda da şu soru ortaya çıkar: Bu bizi nasıl etkileyecek? Yakın dönem YZ ilerlemesi insan olmanın ne anlama geldiğini nasıl etkileyecek? YZ'nin tamamen hedeflere, özgürlüğe, önseziye, yaratıcılığa ya da dile –yani insan olmanın merkezindeki özelliklere– sahip olmadığını söylemenin giderek daha da zor olduğunu gördük. Bu da kısa dönemde bile, YZ'nin bize tüm görevlerde yetişmesinden çok önce, YZ'nin kendimizi nasıl gördüğümüz, YZ tarafından tamamlandığımızda ne yapabileceğimiz ve YZ'ye karşı rekabet ederken para kazanmak için ne yapabileceğimiz üzerinde ciddi etkileri olabilir. Bu etki bizi daha iyiye mi kötüye mi götürecek? Yakın dönemde bize ne gibi zorluklar ve fırsatlar sunacak?

Uygarlık üzerine sevdiğimiz her şey insan zekâsının ürünüdür. Bu yüzden de eğer bunu yapay zekâyla artırırsak, hayatı daha da iyi yapma potansiyeline açık açık sahip olacağız. YZ'deki en basit ilerleme bile bilim ve teknolojide ciddi atılımlara yol açabilir, kaza, hastalık, adaletsizlik, savaş, angarya ve fakirliği düşürebilir. Ancak yeni sorunlar yaratmadan YZ'nin faydalarının hasadını almak için, pek çok önemli soruyu cevaplamamız gerekiyor. Örneğin:

1. Gelecekteki YZ sistemlerini bugüne nazaran nasıl daha sağlam yapabiliriz ki hacklenmeden, çökmeden ve bozulmadan istediğimizi yapabilsin?
2. Hukuk sistemlerimizi daha adil ve etkili yapmak ve devamlı değişen dijital duruma ayak uydurmak için nasıl güncelleyebiliriz?

3. Silahları, ölümcül otonom silahlar üreten kontrolden çıkmış bir silahlanma yarışına sebebiyet vermeden nasıl daha akıllı ve masum, sivilleri öldürme ihtimali daha düşük hale getirebiliriz?
4. Refahımızı otomasyonla, insanları gelirsiz ve amaçsız bırakmadan nasıl artırabiliriz?

Bu bölümün kalan kısmını tüm bu soruları cevaplamaya adayalım. Bu dört yakın dönem soru temelde sırasıyla bilgisayar bilimcilere, hukuk akademisyenlerine, askerî stratejistlere ve ekonomistlere yöneliktir. Ancak cevapları ihtiyacımız olduğu anda alabilmek için, herkesin bu diyaloğa katılması gerekir çünkü göreceğimiz üzere, zorluklar tüm geleneksel sınırları aşıyor – hem uzmanlık alanları arasındaki hem de ulusların arasındaki.

Program Hataları vs. Sağlam YZ

Bilgi teknolojisi bilimden finansa, üretime, ulaşıma, sağlığa, enerjiye ve iletişime kadar insanlık girişimimizin her bölümüne olumlu katkılar sağlamıştır. Bu katkılar, YZ'nin taşıdığı ilerleme potansiyeline kıyasla sönük kalır. Ancak teknolojiye daha çok bel bağladıkça, istediğimiz şeyi yapıyor olması, sağlamlığı ve güvenilirliği daha büyük önem taşımaya başlar.

İnsanlık tarihi boyunca, teknolojimizi faydalı kılmak için aynı test edilip onaylanmış yaklaşıma güvendik: hatalarımızdan öğrenmek. Ateşi keşfettik, defalarca batırdık ve sonrasında da yangın söndürücü, yangın çıkışı, yangın alarmı ve itfaiyeyi geliştirdik. Otomobili icat ettik, defalarca kaza yaptık ve sonrasında emniyet kemeri, hava yastığı ve sürücüsüz arabalar geldi. Şimdiye dek teknolojilerimiz faydalarına kıyasla yeterince az ve sınırlı sayıda kazaya sebebiyet verdi. Karşı durulmaz bir şekilde çok

daha güçlü teknolojiler geliştirirken öyle bir seviyeye ulaştık ki tek bir kazanın bile tüm faydaları geçersiz kılabilecek kadar yıkıcı olabileceği bir noktaya ulaştık. Bazıları kazara çıkacak küresel bir nükleer savaşı buna örnek olarak gösteriyor. Diğerleri de biyomühendislikle ortaya çıkarılabilecek bir salgından bahsediyor. Bir sonraki bölümde, gelecekteki YZ'nin insanlığın soyunun tükenmesine sebebiyet verip vermeyeceğine dair tartışmaları inceleyeceğiz. Fakat hayati bir sonuca ulaşmak için böylesine uç örnekleri düşünmemize gerek yok: Teknoloji giderek daha güçlenirken, güvenlik mühendisliğinde deneme yanılma yöntemine daha da az dayanmalıyız. Diğer bir deyişle, *reaktif değil proaktif olmamız*, kazaların bir kez bile olmasını engellemeyi hedefleyen güvenlik araştırmasına yatırım yapmamız gerekir. Bu yüzden toplumumuz fare kapanı güvenliğinden çok nükleer reaktör güvenliğine daha çok yatırım yapar.

Bölüm 1'de gördüğümüz gibi Porto Riko Konferansı'nda YZ güvenliği araştırmasına yoğun ilgi olmasının sebebi de buydu. Bilgisayarlar ve YZ sistemleri her zaman çökmüştü ama bu sefer farklıydı: YZ yavaş yavaş gerçek dünyaya giriyordu ve eğer elektrik şebekesindeki, borsadaki ya da nükleer silah sistemindeki YZ çökerse bu basit bir sıkıntıdan çok daha fazlası olurdu. Bu bölümün kalan kısmında size şu anki YZ güvenliği tartışmasında hâkim olan ve tüm dünyada takip edilen dört ana teknik YZ güvenlik araştırma alanını tanıtacağım: ***doğrulama***, ***geçerlilik sınaması***, ***güvenlik*** ve ***kontrol***.* İşlerin çok sıkıcı ve çok inek işi olmasını engellemek için, bunu farklı alanlarda bilgi teknolojisinin geçmiş başarı ve başarısızlıkları ile onlardan öğrenebileceğimiz değerli dersler ve yaratabilecekleri araştırma zorluklarını inceleyerek yapalım.

* YZ güvenlik araştırması ortamı hakkında daha detaylı bilgi isterseniz, FLI'dan Richard Mallah'ın başını çektiği tüm camianın emekleriyle geliştirilmiş interaktif bir haritayı buradan bulabilirsiniz: https://futureoflife.org/landscape.

Bu hikâyelerin birçoğu eski, kimsenin YZ diyemeyeceği düşük teknoloji bilgisayar sistemleri hakkında ve neredeyse hiç zayiata sebebiyet vermemiş olsa bile, bize başarısız olmalarının felaketle sonuçlanacağı geleceğin güvenli ve güçlü YZ sistemlerini tasarlamakta değerli dersler vereceğini göreceğiz.

Uzay Keşfi İçin YZ

Sevdiğim bir konuyla başlayalım: uzay keşfi. Bilgisayar teknolojisi Ay'a insan göndermemizi ve Güneş sistemimizdeki tüm gezegenleri keşfetmek için insansız araçlar yollamamızı sağladı, hatta bu araçlar Satürn'ün uydularından biri olan Titan'a ve bir meteora indi bile. Bölüm 6'da göreceğimiz gibi, gelecekteki YZ diğer güneş sistemlerini ve galaksileri keşfetmekte bize yardımcı olabilir. Tabii eğer program hataları olmazsa. 4 Haziran 1996'da, Dünya'nın manyetosferini araştırmayı ümit eden bilim insanları, kendi inşa ettikleri bilimsel araçlar ile Avrupa Uzay Ajansı'nın Ariane 5 isimli roketi uzaya fırlatıldığında büyük bir sevinç duygusuna kapılmışlardı. Otuz yedi saniye sonra, gülümsemeleri kayboldu çünkü roket yüzlerce milyon dolarlık bir havai fişek gösterisi gibi patlamıştı.[7] Sebebi ise kendisi için ayrılan 16 bite sığamayacak kadar büyük bir sayıyı manipüle eden hatalı bir yazılımdı.[8] İki yıl sonra, NASA'nın Mars İklim Uydusu yanlışlıkla kızıl gezegenin atmosferine girerek parçalara ayrıldı çünkü yazılımın iki farklı kısmı kuvvet için farklı birimler kullanarak roket motoru itki kontrolünde %445'lik bir hataya yol açmıştı.[9] Bu NASA'nın ikinci aşırı pahalı hatasıydı: Venüs'e gönderilen Mariner I uzay aracı, uçuş kontrol yazılımının hatalı noktalama işaretinden dolayı bozulmasından sonra 22 Temmuz 1962'de Cape Canaveral'den kalkış yaparken patlamıştı.[10] Uzaya hata gönderme sanatının tek ustalarının Batılılar olmadığını göstermek istercesine, Sovyet

Phobos 1 uçuşu da 2 Eylül 1988'de başarısızlığa uğramıştı. Bu, Mars'ın uydusu Phobos'a iniş yapma hedefindeki, fırlatılmış en ağır gezegenler arası uzay aracıydı ancak Mars yolundayken bir tire eksiği yüzünden "end-of-mission (görev sonu)" mesajı gönderilince tüm sistemler kapanmış ve görev birden sona ermişti.[11]

Bu örneklerden öğrendiğimiz şey bilgisayar bilimcilerinin *doğrulama* dediği şeyin önemidir: Yazılımın tüm beklenen gereklilikleri tamamen yerine getirmesini sağlamak. Tehdit altındaki yaşamlar ve kaynaklar arttıkça, yazılımımızın ilk başta istenildiği şekilde çalışacağına dair daha çok güvence isteriz. Neyse ki YZ, doğrulama sürecinin otomasyonunu yaparak geliştirebilir. Mesela bütünlüklü, genel amaçlı bir işletim sistemi çekirdeği, *seL4*, matematiksel olarak çökmelere ve güvensiz operasyonlara karşı güçlü bir garanti verecek kurallı belirtimlere karşı kontrol edilir: Microsoft Windows ve MacOS'un cazip yanlarına sahip olmasa da "mavi ekran" ya da "kıyametin çıkrığı" isimleriyle bilinen şeyi karşınıza çıkarmayacağına emin olabilirsiniz. ABD Savunma İleri Araştırma Projeleri Ajansı (DARPA) güvenilirliliği kanıtlanabilir olan HACMS (yüksek güvenceli siber askerî sistemler) isimli açık kaynak kodlu yüksek güvenceli bir dizi aracın geliştirilmesine sponsor olmuştu. Önemli bir zorluk da bunun gibi araçları genel kullanıma alınacak kadar güçlü ve kullanımı kolay yapmaktır. Başka bir zorluk da yazılım robotlara ve yeni çevrelere ilerler, geleneksel önden programlanmış yazılımlar devamlı öğrenen, buna bağlı olarak da –Bölüm 2'deki gibi– davranışlarını değiştiren YZ sistemleriyle değiştirilirken doğrulama görevinin giderek daha da zorlaşmasıdır.

Finans İçin YZ

Bilgi teknolojisinin, kaynakların tüm dünyada ışık hızında ve verimli biçimde yeniden dağıtılmasını sağlayarak ve mortgage'dan

startup şirketlerine kadar her konuda bütçeye uygun finansman sunarak dönüştürdüğü bir başka alan da finanstır. YZ'deki ilerleme gelecekte muhtemelen finansal işlemle ilgili çok fazla kâr fırsatı sunacaktır: Çoğu hisse senedi al/sat kararı şu anda bilgisayarlar tarafından otomatik olarak yapılmaktadır ve benim MIT'den mezun olmak üzere olan öğrencilerim algoritmik ticareti geliştirmek için onlara teklif edilen astronomik miktarlara tav oluyorlar.

Doğrulama, finansal yazılım için de önemlidir. Bunu, Amerikan firması Knight Capital 1 Ağustos 2012'de doğrulanmamış bir ticaret yazılımını kullanmaya başladıktan sonraki kırk beş dakika içerisinde 440 milyon dolar kaybederek acı bir şekilde öğrenmişti.[12] 6 Mayıs 2010'daki trilyon dolarlık "Flaş Çökme" başka bir sebepten ötürü önemliydi. Procter & Gamble gibi firmaların hisselerinin bir sentten yüz bin dolara kadar dalgalanmasına sebebiyet vererek, piyasalar normale dönene kadar yarım saat boyunca devasa karmaşa çıkarmış olmasına rağmen sorun, doğrulamanın önleyebileceği program hataları ya da bilgisayar sorunlarından dolayı ortaya çıkmamıştı. Onun yerine, beklentilerin ihlal edilmesinden dolayı gerçekleşmişti: Pek çok farklı şirketin otomatik ticaret programları kendilerini beklenmedik bir durumda çalışırken bulmuşlar, varsayımlarının hiçbirinin geçerli olmadığını görmüşlerdi. Mesela bir borsa bilgisayarı bir hissenin bir sentlik ücreti olduğunu raporluyorsa, o hissenin gerçekten de bir sent değerinde olacağı varsayımı gibi.

Flaş çökme bilgisayar bilimcilerin *geçerlilik sınaması* dediği şeyin önemini gösterir: Doğrulama "Sistemi doğru kurdum mu?" sorusunu sorarken, geçerlilik sınaması "Doğru sistemi kurdum mu?" diye sorar.* Mesela sistem her zaman geçerli olmayabile-

* Daha doğru biçimde ifade etmek gerekirse, doğrulama bir sistemin şartnamelere uyup uymadığını kontrol ederken, geçerlilik sınaması doğru şartnamenin seçilip seçilmediğini kontrol eder.

cek varsayımlar üzerine mi kuruludur? Eğer öyleyse, belirsizliği ortadan kaldırmak için nasıl geliştirilebilir?

Üretim İçin YZ

YZ'nin hem kesinliği hem de verimliliği geliştiren robotları kontrol ederek üretimi iyileştirmek için büyük bir potansiyele sahip olduğunu söylemeye gerek bile yoktur. Giderek gelişen üç boyutlu yazıcılar artık ofis binalarından tuz zerreciğinden küçük mikromekanik cihazlara kadar her şeyin numunelerini yapabiliyor.[14] Devasa endüstriyel robotlar arabalar ve uçaklar yaparken, bütçeye uygun, bilgisayar kontrollü değirmenler, tornalar, kesiciler ve benzeri şeyler yalnızca fabrikaları değil, heveslilerin kendi fikirlerini dünyada binden fazla bulunan "fab lab" isimli üretim laboratuvarlarında gerçekleştirdiği "*maker* hareketi"ni de güçlendiriyor.[15] Fakat etrafımızda ne kadar çok robot bulunursa, yazılımlarını doğrulayıp geçerliliklerini sınamak o kadar çok önem kazanır. Bir robot tarafından öldürülen ilk kişi, Flat Rock, Michigan'daki Ford fabrikasında bir işçi olan Robert Williams'dı. 1979'da, depolama alanından parça getirmesi gereken bir robot arıza yapmıştı ve Williams parçaları kendisi almak için bu alana tırmandı. Robot sessiz biçimde çalışmaya başladı ve kafasını parçaladı. Diğer çalışanlar bunu fark edene kadar robot bu işleme yarım saat kadar devam etmişti.[16] Bir sonraki robot kurbanı Akashi, Japonya'daki Kawasaki fabrikasındaki bakım mühendisi Kenji Urada idi. 1981'de bozuk bir robot üzerinde çalışırken, yanlışlıkla açma düğmesine basmış ve robotun hidrolik kolu altında ezilerek ölmüştü.[17] 2015'te, Baunatal, Almanya'daki Volkswagen üretim tesislerinden birinde yirmi iki yaşındaki bir işçi, robotu otomobil parçalarını almak ve onları manipüle etmek üzerine kurmaya

çalışıyordu. Bir şeyler yanlış gitti ve robot onu yakalayarak metal plakaya çarptı ve öldürdü.[18]



Görsel 3.3: Geleneksel endüstriyel robotlar pahalı ve programlanması zorken, programlama deneyimi olmayan işçilerden ne yapacağını öğrenebilen daha ucuz ve YZ kaynaklı robotları kullanmaya doğru bir eğilim vardır.

Bu kazalar trajik de olsa, tüm endüstriyel kazaların çok ufak bir kısmını teşkil ettiğini söylemek gerek. Dahası, teknoloji ilerledikçe endüstriyel kazalar, ABD'de 1970'teki 14.000 ölümden 2014'te 4821 ölüme düşerek *gerilemiştir* de.[19] Yukarıda bahsedilen üç kaza normalde zekâya sahip olmayan makinelere zekâ eklemenin robotlara insanların etrafında daha dikkatli olmalarını öğreterek endüstriyel güvenliği daha da iyileştirebileceğini gösterir. Üç kaza da daha iyi geçerlilik sınamasıyla önlenebilirdi: Robotlar program hatası ya da kötü niyetle zarar vermiyorlardı, geçersiz varsayımlarda bulundukları için zarar veriyorlardı; kişinin mevcut olmadığı ya da bir otomobil parçası olduğu varsayımıyla.

Ulaşım İçin YZ

YZ üretimde pek çok yaşamı kurtarırken, ulaşımda çok daha fazlasını kurtarma potansiyeline sahiptir. Araba kazaları tek başına

2015 yılında 1.2 milyon ölüme sebebiyet verirken, uçak, tren ve gemi kazaları birlikte binlercesine neden olmuştur. ABD'de, yüksek güvenlik standartlarıyla motosiklet kazaları geçen sene otuz beş bin insanın ölümüne yol açtı; tüm endüstriyel kazaların yedi katı.[20] Bu konuda, 2016 yılında Austin, Texas'ta düzenlenen geleneksel Yapay Zekânın İlerlemesi Birliği toplantısında düzenlenen bir panelde konuştuğumuzda, İsrailli bilgisayar bilimci Moşe Vardi çok duygusallaştı ve YZ'nin yol kazalarını *engelleyebileceğini* değil engellemek *zorunda* olduğunu söyledi: "Bu bir ahlaki zorunluluk!" diye çıkıştı. Neredeyse tüm araba kazaları insan hatasından ötürü gerçekleştiği için, YZ bazlı sürücüsüz arabaların yol kazasına bağlı ölümleri en azından %90 oranında azaltabileceğine inanılıyor ve bu iyimserlik sürücüsüz araçları yollara dökmek için büyük ilerlemeyi ateşliyor. Elon Musk geleceğin sürücüsüz arabalarının yalnızca daha güvenli değil, sahipleri onlara ihtiyaç duymazken onlar için para da kazanabileceğini, hatta Uber ve Lyft'le rekabet edebileceğini söylüyor.

Şu ana kadar sürücüsüz araçların gerçekten de insan sürücülerden daha iyi bir güvenlik karnesi var ve gerçekleşen kazalar geçerlilik sınamasının önemi ve zorluğunun altını çiziyor. Google'ın sürücüsüz aracının yol açtığı ilk küçük kaza 14 Şubat 2016'da gerçekleşmişti. Kazanın sebebi bir otobüs hakkında yapılan yanlış varsayımdı: araba önüne çıkınca otobüs sürücüsünün duracağı varsayımı. Sürücüsüz bir Tesla'nın bir kamyonun römorkuna çarpmasıyla oluşan ilk ölümcül kaza iki kötü varsayıma dayanıyordu:[21] römorkun parlak beyaz tarafının parlak gökyüzünün bir parçası olduğu ve şoförün (söylenene göre Harry Potter filmi izliyordu) dikkat ettiği ve yanlış bir şey olursa müdahale edeceği varsayımlarına.*

* Bu kazayla bile, istatistiklere göre Tesla'nın oto pilotu kazaları yüzde kırk oranında azaltmaktadır: http://tinyurl.com/teslasafety.

Fakat bazen iyi doğrulama ve geçerlilik sınaması da kazalardan kaçınmak için yeterli değildir çünkü ayrıca iyi *kontrole* de ihtiyacımız vardır: bir insan operatörün sistemi izleyebilme ve gerekli olduğu takdirde davranışını değiştirebilme yetisi. Böyle *insanın içinde olduğu* sistemlerin iyi çalışması için, insan makine iletişiminin etkili olması hayati önem taşır. Tıpkı paneldeki kırmızı ışığın yanlışlıkla bagajınızı açık bıraktığınızı söylemesi gibi. Aksine, İngiliz arabalı vapuru *Herald of Free Enterprise* 6 Mart 1987'de Zeebrugge Limanı'nı kapıları açık biçimde terk ettiğinde, kaptan için herhangi bir görülebilir uyarı yoktu. Vapur limandan ayrıldıktan kısa süre sonra alabora olarak yüz doksan üç kişinin ölümüne yol açmıştı.[22]

Daha sağlıklı insan ve makine iletişimiyle önlenebilecek bir diğer trajik kontrol hatası 1 Haziran 2009 gecesi, 447 uçuş numaralı Air France uçağı Atlantik Okyanusu'na çakılarak iki yüz yirmi sekiz kişinin ölümüne sebep olduğunda gerçekleşti. Resmî kaza raporuna göre, her şey için çok geç olana dek "mürettebat motorun durduğunu anlamamış ve uçağın burnunu aşağı itmek gibi toparlanma manevralarını asla uygulamamıştı". Uçuş güvenliği uzmanları kokpitte pilotlara burnun çok fazla yukarı kalkık olduğunu gösteren bir "hücum açısı" göstergesi olsaydı kazanın engellenebileceğini söylüyorlardı.[23]

20 Ocak 1992'de 148 uçuş numaralı Air Inter Strasbourg uçağı, Fransa yakınlarındaki Vosges Dağları'na çarpıp seksen yedi kişinin ölümüne sebep olduğunda, kazanın sebebi insan ve makine iletişimi değil, kafa karıştırıcı bir kullanıcı arayüzüydü. Pilotlar küçük klavyeye "33" sayısını girdiler çünkü 3,3 açılık bir alçalma açısı istiyorlardı ancak oto pilot bunu dakikada 3300 feet olarak algıladı çünkü farklı bir çalışma modundaydı ve görüntüleme ekranı modu gösteremeyecek kadar küçük olduğu için pilotlar hatalarının farkına varamadılar.

Enerji İçin YZ

Bilgi teknolojisi dünyanın tüm elektrik şebekelerinde üretim ile tüketimi dengeleyecek sofistike algoritmalar ve enerji santrallerini güvenli ve verimli biçimde çalıştıracak sofistike kontrol sistemleriyle enerji üretimi ve dağıtımı için harikalar yaratmıştı. Gelecekteki YZ ilerlemesi, "akıllı şebekeyi" daha da akıllı yaparak bireysel çatı güneş panelleri ve ev pil sistemleri seviyesine kadar değişen arz ve talebe uyum sağlayacak hale getirebilir. Fakat 14 Ağustos 2003 Perşembe günü, ABD ve Kanada'da elli beş milyona yakın insan için elektrikler gitti ve pek çoğu günler boyunca elektriksiz kaldı. Burada da ana sebep başarısız olmuş makine-insan iletişimiydi: Bir yazılım hatası, Ohio'daki bir kontrol odasındaki alarm sisteminin küçük bir problem (aşırı yüklenmiş iletim hatlarının budanmamış yeşilliğe çarpması) kontrolden çıkmadan önce enerjiyi yeniden dağıtma ihtiyacı doğduğu konusunda operatörleri uyarmasını engellemişti.[24]

Three Mile Adası, Pennsylvania'da 28 Mart 1979'da bir reaktörde gerçekleşen kısmi nükleer erime, milyar dolarlık temizleme maliyetine ve nükleer enerjiye karşı sert bir muhalefete yol açmıştı. Son kaza raporu kötü kullanıcı arayüzünün yol açtığı bir karışıklık da dâhil olmak üzere pek çok etken faktörü listelemişti.[25] Özel olarak, operatörlerin güvenlik tehlikesi taşıyan bir valfin açık ya da kapalı olduğunu gösterdiğini düşündükleri uyarı ışığı, valfi kapatmak için bir sinyalin gönderilip gönderilmediğini gösteriyordu. Bu yüzden de operatörler valfin açık kaldığını fark etmediler.

Bu enerji ve ulaşım kazaları, gün geçtikçe daha çok fiziksel sistemi YZ'nin eline bırakırken, yalnızca makinelerin kendi başlarına çalışmaları için değil, insan kontrolörle etkili biçimde işbirliği yapmaları için de ciddi araştırmalar gerektiğini öğretiyor. YZ zekileştikçe, bu yalnızca bilgi paylaşımı için iyi kulla-

nıcı arayüzleri inşa etmeyi değil, insan ve bilgisayar ekiplerine optimal biçimde görev paylaştırmayı nasıl yapmamız gerektiğini öğrenmeyi de kapsayacak. Mesela, kontrolün transfer edilmesi gereken durumları tespit etmek ve yüksek değerli tercihler için gereksiz bilgi akışıyla insan kontrolörlerin dikkatini dağıtmak yerine insan muhakemesini uygulamak gibi.

Sağlık Hizmetleri İçin YZ

YZ'nin sağlık hizmetlerini iyileştirmek için de yüksek potansiyeli vardır. Medikal kayıtların dijitalleştirilmesi, doktor ve hastaların daha hızlı ve daha iyi tercihler yapmasını ve dijital görüntülerin teşhis edilmesinde dünyanın dört bir yanındaki uzmanlardan anında yardım almalarını çoktan sağladı. Hakikaten, bilgisayar görüntüleme ve derin öğrenmedeki gelişmeyi düşününce böyle teşhislerin yapılması için en iyi uzmanlar yakında YZ sistemleri olabilir. Örneğin, 2015'te Hollanda'da yapılan bir çalışma manyetik rezonans görüntüleme (MRI) kullanılarak yapılan prostat kanseri bilgisayar teşhisinin insan radyologlarınki kadar iyi olduğunu[26] ve 2016'da Stanford'da yapılan bir çalışma da YZ'nin mikroskop görüntüleri kullanarak akciğer kanserini insan patologlardan daha iyi tespit edebildiğini göstermişti.[27] Eğer makine öğrenimi genler, hastalıklar ve tedaviye verilen cevaplar arasındaki ilişkileri açığa çıkarmaya yardımcı olabilirse, kişiselleşmiş tıpta devrime yol açabilir, çiftlik hayvanlarını daha sağlıklı yapabilir ve daha dayanıklı bitkiler oluşturabilir. Dahası, robotların gelişmiş bir YZ bile kullanmadan insanlardan daha kesin ve güvenilir cerrahlar olma potansiyeli vardır. Geniş bir çeşitlilikte robotik ameliyatlar son yıllarda başarılı bir şekilde gerçekleştirilmiş, daha az kan kaybı, daha az ağrı ve kısa iyileşme süresine yol açan daha ufak kesikler, kesinlik ve daha küçük cihazları da mümkün kılmıştır.

Maalesef, sağlık endüstrisinde de sağlam yazılımların önemine dair acı verici dersler alındı. Mesela, Kanada yapımı Therac-25 radyasyon terapi cihazı, kanser hastalarını iki farklı modda tedavi etmek için tasarlanmıştı: ya düşük güçlü elektron demeti ya da hedef üzerinde özel bir kalkanla tutulan yüksek güçlü megavolt X-ray. Maalesef doğrulanmamış hatalı yazılım, bazen teknisyenlerin düşük güçlü demeti verdiklerini sanarak kalkansız biçimde megavolt ışını vermelerine yol açmış ve bu birkaç hastanın yaşamına mal olmuştu.[28] Çok daha fazla hasta, geçerliliği doğru şekilde sınanmamış kafa karıştırıcı bir kullanıcı arayüzü sebebiyle 2000 ve 2001'de radyoaktif kobalt-60 kullanan radyoterapi ekipmanının aşırı maruz kalma zamanlarına programlanmış olduğu Panama'daki Ulusal Onkoloji Enstitüsü'nde radyasyon doz aşımından hayatını kaybetmişti.[29] Yakın zamanlarda yayınlanan bir rapora göre,[30] robotik ameliyat kazaları ABD'de 2000 ile 2013 arasında, yalnızca elektrik atlama, hastanın üzerine düşen yanık ya da kırık parçalar gibi donanım sorunlarına ek olarak kontrol edilemeyen hareketler ve spontane gelişen enerji kesintileri gibi yazılım problemlerini içeren sık görülen sorunlar ile 144 ölüm ve 1391 yaralanmayla ilişkilendirilmiştir.

Fakat raporun incelediği yaklaşık iki milyon robotik ameliyatın kalanı sıkıntısız biçimde geçmiştir ve robotlar ameliyatı daha güvenli hale getirmektedirler. Bir ABD hükümet araştırmasına göre, kötü hastane bakımı tek başına ABD'de yılda yüz binden fazla ölüme yol açmaktadır.[31] Bu yüzden de tıp için daha iyi YZ geliştirmek için var olan ahlaki sorumluluk sürücüsüz arabalarda olduğundan daha bile güçlüdür.

İletişim İçin YZ

İletişim endüstrisi belki de bilgisayarların şu ana kadar en büyük etkiyi yarattığı alandır. Ellilerde bilgisayar donanımlı tele-

fon santralinin, altmışlarda internetin ve 1989'da World Wide Web'in piyasaya sürülmesinden sonra, milyarlarca insan iletişim kurmak, alışveriş yapmak, haberleri okumak, film izlemek ya da oyun oynamak için çevrimiçi oldular ve dünyanın tüm bilgisine tek bir tıkla ulaşmaya da alıştılar, genelde de bedavaya. Gelişen, *şeylerin interneti* lambalar, termostatlar ve dondurucular dan çiftlik hayvanları üzerindeki biyoçip vericilere kadar her şeyi çevrimiçi yaparak verimlilik, hassasiyet, elverişlilik ve ekonomik uygunluk konularında iyileştirme vaadinde bulunuyor.

Dünyayı birbirine bağlamaktaki bu görkemli başarılar, bilgisayar bilimcilere dördüncü bir zorluk daha getirdi: Yalnızca doğrulama, geçerlilik sınaması ve kontrolü değil, kötü niyetli yazılımlara ("malware") ve hacklere karşı *güvenliği* de iyileştirmeleri gerekiyor. Daha önce bahsettiğim sorunların her biri istemsiz hatalardan kaynaklanmış olsa da güvenlik, *kasti suistimali* hedef alır. Ciddi medya dikkatini üstüne çeken ilk kötü niyetli yazılım UNIX işletim sistemlerindeki yazılım hatalarını istismar eden, 2 Kasım 1988'de ortaya çıkan "Morris solucanı"ydı. Çevrimiçi kaç tane bilgisayar olduğunu saymak için yapılmış ve yanlış yönlere ilerlemiş bir girişim olduğu söyleniyordu ve o zaman interneti oluşturan altmış bin bilgisayarın yüzde onuna girip çökertmiş olsa da yaratıcısı Robert Morris'in MIT'de bilgisayar bilimi alanında kadrolu bir profesörlük almasını engellemedi.

Bir de yazılımı değil insanların zaaflarını istismar eden kötü amaçlı yazılımlar vardır. Sanki benim doğum günümü kutlamak için yapılmış gibi 5 Mayıs 2000'de insanlar, tanıdıkları ve meslektaşlarından "ILOVEYOU (SENİSEVİYORUM)" başlıklı e-postalar aldılar ve ekteki "LOVE-LETTER-FOR-YOU.txt.vbs (SENİN-İÇİN-AŞK-MEKTUBU.txt.vbs)" dosyasına tıklayan Microsoft Windows kullanıcıları farkında olmayarak bilgisayarlarına zarar veren ve bu e-postayı adres defterlerindeki herkese gönderen bir

komut dizisini çalıştırmış oldu. Filipinler'deki iki genç yazılımcı tarafından üretilen bu solucan tıpkı Morris'in solucanı gibi internetin yaklaşık yüzde onunu etkiledi ama internet o zamanlar çok daha büyük olduğu için elli milyon bilgisayarı etkileyen ve beş milyar doların üzerinde hasara yol açan tüm zamanların en büyük salgınlarından biri oldu. Acılı bir biçimde farkında olduğunuz üzere, internet halen daha uzmanların solucanlar, Trojanlar, virüsler ve diğer rahatsız edici isimli kategorilere ayırdığı pek çok bulaşıcı kötü niyetli yazılımla doludur ve yol açtıkları hasar, zararsız şaka mesajları göstermekten dosyalarınızı silmeye, kişisel bilgilerinizi çalmaya, sizi gözetlemeye ve bilgisayarınızı spam göndermek için kullanmaya kadar uzanır.

Kötü niyetli yazılımlar ulaşabildikleri her bilgisayarı hedef alırken, *hackerlar* özel hedeflere saldırır. Son zamanlarda gördüğümüz yüksek profilli örnekler arasında Target, TJ Maxx, Sony Pictures, Ashley Madison, Suudi petrol şirketi Aramco ve ABD Demokrat Ulusal Komitesi bulunuyor. Dahası, bu saldırılar giderek daha da şaşırtıcı olmaya başlıyor. Hackerlar 2008 yılında Heartland Ödeme Sistemleri'nden yüz otuz milyon kredi kartı numarası ve diğer hesap bilgilerini çaldı ve 2013 yılında bir milyardan (!) fazla Yahoo! e-posta hesabına girdiler.[32] ABD hükümetinin Personel Yönetimi Ofisi'ne 2014 yılında yapılan bir saldırı, söylendiğine göre üst düzey güvenlik iznine sahip çalışanlar ve gizli ajanların parmak izleri de dâhil olmak üzere, yirmi bir milyondan fazla insana ait personel kayıtları ve iş başvuru bilgilerini sızdırmıştı.

Sonuç olarak yeni bir sistemin %100 güvenli ve hacklenemez olduğu iddiasını okuduğumda hep gözlerimi deviririm. Yine de "hacklenemez" kelimesi gelecekteki YZ sistemlerinin, onları hayati önem taşıyan altyapıların ya da silah sistemlerinin başına geçirmeden önce nasıl olması gerektiğini özetler. Yani YZ'nin toplumda artan rolü bilgisayar güvenliği için işi giderek ciddiye

bindirmektedir. Bazı hackler insanın aptallığını ya da yeni yayınlanmış bir yazılımın karmaşık zayıflıklarını istismar ederken, bazıları da utanç verici bir biçimde uzun bir süre fark edilmeden gitmiş basit program hatalarını kullanarak uzak bilgisayarlara yetkilendirilmemiş giriş sağlar. "Heartbleed" hatası bilgisayarlar arası güvenli iletişim için kullanılan en popüler yazılım kütüphanelerinden birinde 2012'den 2014'e kadar durmuştur. Aynı şekilde, "Bashdoor" hatası 1989'dan 2014'e kadar Unix bilgisayarlarının işletim sisteminde gömülü kalmıştır. Bu, YZ araçları için gelişmiş doğrulama ve geçerlilik sınamasının güvenliği de artıracağı anlamına gelir.

Maalesef daha iyi YZ sistemleri yeni zayıflıklar bulmada ve daha sofistike hackler düzenlemekte de kullanılabilir. Mesela, bir gün kişisel bilgi vermeniz için sizi ikna etmeye çalışan, ilginç biçimde kişiselleşmiş bir e-posta "oltalaması" aldığınızı düşünün. Bir YZ tarafından arkadaşınızın hacklenmiş hesabından gönderilmiş ve onu taklit ediyor, diğer postalarının analizini yaptığı için de yazı tarzını taklit ediyor ve diğer kaynaklardan çektiği birçok kişisel bilginizi de kullanıyor. Bunu yer misiniz? Peki ya bu e-posta doğrudan sizin kredi kartı şirketinizden gelse ve YZ yapımı olduğunu tahmin edemeyeceğiniz dostça bir insan sesiyle yapılmış bir telefon görüşmesi bunu takip etse? Saldırı ile savunma arasında devam eden bilgisayar güvenliği yarışında, savunmanın kazandığını gösteren pek az kanıt vardır.

Hukuk

İşbirliği yapabilme yetimiz sayesinde geri kalan tüm türleri boyunduruk altına alıp dünyayı fethetmiş olan sosyal hayvanlarız biz. İşbirliğini kolaylaştırmak ve teşvik etmek için yasalar geliştirdik. Eğer YZ hukuki ve yönetsel sistemlerimizi geliştirebilirse,

içimizdeki cevheri çıkararak her zamankinden daha iyi işbirliği yapmamızı sağlayabilir. Bu alanda, hem kanunların yazılışı hem de uygulanışında, çok fazla gelişim ihtimali mevcut. O halde sırayla bunları inceleyelim.

Ülkenizdeki adalet sistemini düşündüğünüzde aklınıza gelen ilk şey nedir? Uzun ertelemeler, yüksek maliyetler ve ara sıra yaşanan adaletsizliklerse, yalnız değilsiniz. İlk aklınıza gelen şeyler "adil" ya da "verimli" olsaydı muhteşem olmaz mıydı? Hukuki süreçler soyut olarak bir hesaplama olarak görülebilir, kanıtı ve yasaları girdi olarak alır ve bir karar çıktısı verir. Bu yüzden de bazı bilim insanları bu süreci tamamen *robohâkimler* ile otomatikleştirmeyi hayal ediyor: önyargı, yorgunluk, en son bilgiyi haiz olmama gibi insan hatalarına düşmeyerek her karara yorulmadan aynı yüksek hukuki standartları uygulayan YZ sistemleri.

Robohâkimler

Byron De La Beckwith Jr. 1994 yılında insan hakları sözcüsü Medgar Evers'ı 1963 yılında öldürmekten dolayı hüküm gitmişti. Ancak cinayetten bir yıl sonra, aynı fiziki kanıtlar mevcutken iki ayrı tamamen beyazlardan oluşan Mississippi jürisi bu kararı verememişti.[33] Ne yazık ki hukuk tarihi deri rengi, cinsiyet, cinsel yönelim, din, millet ve diğer faktörlerden dolayı önyargıyla verilmiş kararlarla doludur. Prensipte robohâkimler tarihte ilk kez herkesin yasalar önünde tamamen eşit olmasını sağlayabilir: Her biri aynı olacak ve herkese eşit davranacak, hukuku tamamıyla önyargısız ve şeffaf bir biçimde uygulayacak şekilde programlanabilirler.

Robohâkimler kazaen olan insan önyargılarını da ortadan kaldırabilecekler. Mesela, 2012 yılında İsrailli hâkimler üzerine yapılan tartışmalı bir araştırma hâkimlerin aç olduklarında önemli ölçüde sert kararlar verdiklerini iddia etmişti: Kahvaltıdan sonra

%35'i şartlı tahliyeyi reddederken, öğle yemeğinden önce %85'i reddediyordu.[34] İnsan hâkimlerin diğer bir eksikliği de bir vakanın tüm detaylarını incelemek için yeterli vakitlerinin olmayışıdır. Tam aksine, robohâkimler bir yazılımdan biraz daha fazlası oldukları için tüm bekleyen davaları seri halinde değil ama eş zamanlı görmek için kolayca kopyalanabilir ve her dava gerektiği süre boyunca kendi robohâkimine sahip olabilir. Son olarak, insan hâkimlerin çetrefilli patent tartışmalarından son teknoloji adli tıp tekniklerine dayalı cinayetlere kadar pek çok konuda yeterince teknik bilgiye sahip olmaları mümkün değilken, geleceğin robohâkimlerinin sınırsız bellek ve öğrenme kapasiteleri olabilir.

Önyargısız, yetkin ve şeffaf oldukları için bir gün böyle robohâkimler hem daha etkili hem de daha adil olabilirler. Verimlilikleri de onları daha adil yapar: Hukuki süreci hızlandırarak ve deneyimli avukatların sonucu çarpıtmasını güçleştirerek, mahkemelerde adaletin sağlanmasını çok daha ucuzlaştırabilirler. Bu da avukatlar ordusuna sahip bir milyardere ya da çok uluslu bir şirkete karşı parasız bir bireyin ya da bir *startup* firmasının kazanma şansını artırır.

Öte yandan, robohâkimlerin program hataları olursa ya da hacklenirlerse ne olur? Otomatik oylama makineleri bu ikisinden de çok çekmişti ve parmaklıklar ardında yıllar ya da bankadaki milyonlar söz konusu olduğunda, siber saldırı ihtimali çok daha artar. YZ, bir robohâkimin yasal algoritmayı kullandığına güvenebileceğimiz kadar sağlam yapılmış olsa bile, kararına saygı duyacak kadar mantıksal çıkarımını anladığımızı hissedecek miyiz? Bu zorluk anlaşılmazlık pahasına anlaması kolay geleneksel YZ algoritmalarından daha iyi performans gösteren sinirsel ağlar üzerine yapılmış son çalışmalarla daha da şiddetlenmişti. Eğer sanıklar *neden* hüküm giydiklerini bilmek isterlerse, "Sistemi pek çok veriyle eğittik ve verdiği karar bu," gibi bir cevaptan daha

iyisini hak etmezler mi? Dahası, son araştırmalar derin sinirsel bir öğrenme sistemini yüksek miktarda suçlu verisiyle beslerseniz, kimin suç işlemeye geri dönme ihtimalinin daha yüksek olduğunu (ve dolayısıyla şartlı tahliye verilmemesi gerektiğini) insan yargıçlardan daha iyi tahmin ettiğini göstermişti. Peki, ama sistem suça geri dönmenin istatistiki olarak bir suçlunun cinsiyeti ya da ırkına bağlı olduğunu bulursa – bu durum yeniden programlanmaya ihtiyaç duyan seksist ve ırkçı bir robohâkimin göstergesi midir? Gerçekten de 2016 yılında yapılan bir çalışma ABD'de kullanılan suça geri dönme tahmin yazılımlarının Afrikalı-Amerikalılara karşı önyargılı olduğunu ve adil olmayan cezalara katkı sunduğunu göstermişti.[35] YZ'nin faydalı olmayı sürdürmesi için ölçüp tartmamız gereken önemli sorular bunlar. Robohâkimlerle ilgili ya hep ya hiç tarzı bir karar vermek zorunda değiliz, bu karar daha çok hukuki sistemimizde YZ'yi kullanmak istediğimiz yaygınlık ve hıza dair bir karar. İnsan hâkimlerin, tıpkı yarının doktorları gibi, YZ bazlı karar destek sistemleri olmasını istiyor muyuz? Daha ileri gitmek ve insan hâkimlere temyize götürülebilecek robohâkim kararları mı istiyoruz, yoksa ölüm cezalarında bile son sözü makinelere bırakmak mı?

Hukuki İhtilaflar

Şu ana kadar yalnızca hukukun *uygulanmasını* inceledik; şimdi de *içeriğine* bakalım. Yasalarımızın teknolojiye ayak uydurması için değişmesi gerektiğine dair geniş bir fikir birliği mevcut. Mesela, daha önceden bahsettiğimiz SENİSEVİYORUM solucanını yaratan ve milyarlarca dolarlık hasara sebebiyet veren iki programcı tüm suçlamalardan beraat edip özgür kalmışlardı çünkü o sıralarda Filipinler'de kötü amaçlı yazılımlara karşı hiçbir yasa yoktu. Teknolojik ilerlemenin ivmesi arttığı için, yasaların çok

daha hızlı biçimde güncellenmesi gerekiyor çünkü geride kalma olasılıkları çok daha yüksek. Hukuk okullarına ve hükümetlere daha fazla teknolojiyle içli dışlı insanları almak toplum için zekice bir hamle olur. Fakat oy kullananlar ve kanun yapanlar için YZ bazlı karar destek sistemlerini robot kanun yapıcılar da izlemeli mi?

YZ gelişmelerini yansıtmaları için yasalarımızı en iyi nasıl değiştirmemiz gerektiğine dair tartışmalar baş döndürücü bir biçimde ihtilaflıdır. Bir tartışma, mahremiyet ve bilginin özgürlüğü arasındaki gerilimi yansıtır. Özgürlük taraftarları ne kadar az mahremiyete sahip olursak, mahkemelerin o kadar çok kanıta sahip olacağını ve daha adil kararlar vereceğini ileri sürüyor. Mesela, devlet herkesin elektronik cihazlarını dinler, nerede olduklarını, ne yazdıklarını, neye tıkladıklarını, ne söyleyip yaptıklarını kaydederse, pek çok suç baştan çözülmüş olur ve yenilerinin olması da engellenebilir. Mahremiyet taraftarları ise Orwellci bir gözetim devleti istemediklerini, isteseler bile bunun inanılmaz ölçüde totaliter bir diktatörlüğe dönüşme riski olduğunu söylüyorlar. Dahası makine öğrenimi teknikleri fMRI tarayıcılarından aldıkları beyin verisini, kişinin ne hakkında düşündüğünü, doğruyu söyleyip söylemediğini tespit etmek amacıyla analiz etme konusunda giderek daha da iyileşiyor.[36] Eğer YZ destekli beyin görüntüleme teknolojisi mahkemelerde kullanılmaya başlarsa, oldukça usandırıcı bir süreç olan bir davanın hakikatlerini ortaya koyma süreci çok daha basitleşir ve hızlanır, bu da daha hızlı davalara ve daha adil kararlara yol açar. Ancak mahremiyet taraftarları böyle bir sistemin hata yapıp yapmayacağı ve daha da temel olarak, zihinlerimizin hükümetin burnunu sokabileceği bir yer olup olmaması gerektiği konularında endişe ediyorlar. Adalet ve mahremiyet, toplumu korumak ile kişisel özgürlüğü korumak arasında çizgiyi *siz* nereden çekerdiniz? Nereden çekerseniz çekin, kanıtların taklit edilmesi kolaylaştıkça, yavaş ama kararlı

bir biçimde giderek azalan bir mahremiyete doğru ilerleyecek mi? Mesela, YZ sizi suç işlerken gösterebilecek tamamen gerçekçi sahte videolar üretebilmeye başladığı anda, devletin daima herkesin nerede olduğunu bildiği ve gerektiğinde size gerçekliği reddedilemez bir kanıt sunabildiği bir sisteme oy verir miydiniz?

Başka bir ciddi tartışma da YZ araştırmasının denetim altına alınıp alınmaması üzerinedir. Daha genel olarak ise, faydalı bir sonucun elde edilme ihtimalini maksimize etmek için politikacıların YZ araştırmacılarına ne gibi teşvikler vermesi gerektiği üzerine tartışmalar var diyelim. Bazı YZ araştırmacıları YZ gelişiminde her türlü denetime karşı olmuşlardı. Çünkü bunun acilen gereken yeniliği (mesela, hayat kurtaran sürücüsüz arabalar) gereksiz yere erteleyeceğini ve ileri derece YZ araştırmasını yeraltına ve/veya daha hoşgörülü hükümetleri olan diğer ülkelere iteceğini düşünüyorlardı. İlk bölümde bahsettiğimiz Porto Riko faydalı YZ konferansında, Elon Musk hükümetlerden şu anda beklediğimiz şeyin idare değil iç görü olduğunu söylemişti: YZ'nin gelişimini izleyebilecek ve gerekirse yönlendirebilecek teknik açıdan yetkin insanların devlet kademelerinde bulunmasıydı. Devlet regülasyonlarının bazen süreci durdurmaktan çok besleyebildiğini de söyledi: Mesela, eğer sürücüsüz araçlar için devlet güvenlik standartları sürücüsüz araç kazalarının sayısını azaltmakta yardımcı olabilirse, toplumsal tepki daha az olur ve yeni teknolojinin benimsenmesi hızlanabilir. Güvenliğe en çok önem veren YZ şirketleri de bu sebeple, kendi yüksek güvenlik standartlarını yakalaması için daha az titiz rakiplerini zorlayacak regülasyonları destekleyebilir.

Başka bir hukuki tartışma da makinelere hakların tanınmasını içeriyor. Eğer sürücüsüz araçlar ABD'de gerçekleşen yıllık otuz iki bin trafik kazasına bağlı ölümleri yarıya indirebilirse, araba üreticileri on altı bin teşekkür mektubu almayacak ama on altı bin davayla karşılaşacaklar. Peki bir sürücüsüz araç kazaya

sebebiyet verirse, kim suçlu olacak – arabadakiler mi, sahibi mi yoksa üreticisi mi? Hukukçu David Vladeck dördüncü bir cevap sunuyor: arabanın kendisi! Özel olarak da sürücüsüz araçların araç sigortasına sahip olmasına izin verilmesini (hatta mecbur tutulmasını) öneriyor. Böylece, hakiki güvenlik geçmişi olan modeller insan sürücülerden daha düşük, oldukça düşük primlere sahip olabilecekken, dikkatsiz üreticilerin elinden çıkmış kötü tasarlanmış modeller yalnızca çok pahalı sigorta poliçelerine sahip olabilecek.

Fakat arabalar gibi makineler de sigorta poliçesine sahip olabilecekse, para ya da mülke sahip olabilmeleri de gerekmez mi? Eğer öyleyse, akıllı bilgisayarların borsada para kazanmasını ve çevrimiçi hizmetler almakta bu parayı kullanmasını önleyecek hukuki hiçbir dayanak kalmaz. Bilgisayarlar insanlara çalışmaları için para ödemeye başladığı anda, insanların yapabildiği her şeyi yapabilir hale gelir. Eğer YZ sistemleri yatırım konusunda insanlardan daha iyi hale gelirse (ki zaten bazı alanlarda öyleler) bu, ekonomimizin çoğunun makinelerce sahiplenildiği ve kontrol edildiği bir duruma yol açacaktır. İstediğimiz şey bu mu? Bu size çok uçuk geliyorsa, ekonomimizin büyük bir kısmının zaten başka bir tüzel kişilik tarafından sahiplenildiğini hatırlayın: genelde içlerindeki herhangi bir insandan daha güçlü olan ve kendilerine ait bir yaşamları da olabilen şirketler.

Makinelere mülk edinme hakkı verilmesi konusunda probleminiz yoksa, oy hakkı konusunda ne düşünürsünüz? Bununla da sorununuz yok diyelim, yeterince zenginse kendinin trilyon tane bulut kopyasını yaparak her seçimin sonucunu belirleyebilecek olmasına rağmen her bilgisayar programının bir oy hakkı mı olmalı? Olmamalıysa, hangi ahlaki temeli baz alarak makine zihnini insan zihninden ayırıyoruz? Makine zihinlerinin bizimkiler gibi öznel deneyimlere sahip olmak açısından bilinçli olmaları bir

fark yaratır mı? Bilgisayarların dünyamız üzerindeki kontrolüyle ilgili bu ihtilaflı soruları bir sonraki bölümde, makine bilinciyle ilgili olanları da Bölüm 8'de irdeleyeceğiz.

Silahlar

Kadim zamanlardan beri insanlık açlık, hastalık ve savaştan çok çekmiştir. YZ'nin açlık ve hastalıkları azaltmakta bize nasıl yardım edebileceğinden zaten bahsettik, peki ya savaş? Bazıları nükleer silahların, onlara sahip olan ülkeler arasında savaştan caydırıcı olduğunu çünkü bu silahların korku verici olduğunu ileri sürüyorlar. Peki, savaşı tamamıyla ortadan kaldırma ümidiyle tüm ulusların çok daha korkutucu, YZ bazlı silahlar inşa etmesi nasıl olurdu? Eğer bu argüman sizi ikna etmediyse ve gelecekteki savaşların kaçınılmaz olacağına inanıyorsanız, bu savaşları daha insancıl yapmak için YZ kullanmaya ne dersiniz? Eğer savaşlar yalnızca makinelerin birbiriyle savaşmasından ibaret olursa, hiçbir insan asker ve sivil ölmek zorunda değil. Dahası, geleceğin YZ destekli *drone*ları ve diğer otonom silah sistemleri (OSS; düşmanları tarafından "katil robotlar" olarak da bilinir) insan askerlerden daha adil ve rasyonel de olabilirler: İnsanüstü sensörlerle donanmış, ölmekten korkmayan robotlar sakin kalabilir, savaşın sıcaklığında bile dengeli ve hesaplı hareket edebilirler ve sivilleri yanlışlıkla öldürme ihtimalleri daha düşük olabilir.

Döngüde Bir İnsan

Peki ya bu otomatik sistemlerde hata olursa, karışıklık çıkarsa ya da beklendiği gibi davranmazlarsa? Aegis sınıfı kruvazörler için geliştirilen ABD Phalanx sistemi, gemisavar füzeleri ve uçaklar gibi tehlikeleri otomatik biçimde tespit ediyor, takip ediyor ve onlara saldırıyor. USS *Vincennes*, Aegis sistemi yüzünden Robokruvazör



Görsel 3.4: Bu ABD Hava Kuvvetleri MQ-1 Predatör droneu gibi günümüzün askerî droneları insanlar tarafından uzaktan kumanda edilse de gelecekteki YZ'yle yönetilen dronelar insanları döngüden çıkartabilir ve bir algoritmayla kimi hedefleyip öldürebileceğini seçebilir.

denilen bir güdümlü füze kruvazörüydü ve 3 Temmuz 1988'de, İran-Irak Savaşı sırasında İran gambotlarıyla giriştiği çarpışmada radar sistemi yaklaşan bir uçağın uyarısını verdi. Kaptan William Rodgers III dalışa geçen bir İran F-14 savaş uçağıyla karşı karşıya olduklarını anladı ve Aegis sistemine ateş etme izni verdi. O sırada farkında olmadığı ise, düşürdüğü uçağın sivil bir İran yolcu jeti olan 655 uçuş numaralı Iran Air uçağı olduğuydu. Bu saldırı uçaktaki iki yüz doksan kişiyi de öldürmüştü ve uluslararası bir tepkiye yol açmıştı. Takip eden süreçte yapılan soruşturmalar radar ekranındaki hangi noktaların sivil uçaklar (655 her zamanki günlük uçuş rotasını izliyordu ve sivil uçak alıcısı açıktı) olduğunu ya da hangi noktaların alçalıp (saldırıya geçmek için) hangilerinin yükseldiğini (Tahran'dan kalkan 655 gibi) otomatik olarak göstermeyen kafa karıştırıcı bir kullanıcı arayüzünü işaret ediyordu. Bunun yerine otomatik sistem, gizemli uçak hakkında bilgi almak için sorgulandığında, "alçalan" bir uçak olduğunu söylemişti çünkü donanma tarafından uçakları

izlemek için kullanılan bir sayıya yanlışlıkla yeniden atadığı bir başka uçağın durumu buydu: Alçalışa geçen uçak aslında çok uzaklarda Umman Körfezi'nde görev yapan bir ABD su üstü savaş hava devriye uçağıydı.

Bu örnekte, zaman baskısından ötürü otomatik sistemin ona söylediğine çok fazla güvenmek zorunda kalmış, son kararı veren bir insan döngünün içindeydi. Şimdiye dek, tüm dünyadan savunma yetkililerine göre, tüm aktif silah sistemlerinin döngüsünde bir insan bulunur. Bunun tek istisnası kara mayınları gibi düşük teknolojili bubi tuzaklarıdır. Ancak şu anda kendi başlarına seçim yapıp hedeflere saldıran, tamamıyla otonom silahlar geliştiriliyor. Tüm insanları döngüden çıkararak hız kazanmak askerî amaçlar açısından çekici: Hemen tepki verebilen tamamen otonom bir dronela dünyanın öte yakasındaki bir insan tarafından uzaktan kontrol edildiği için daha yavaş tepki veren bir drone arasındaki it dalaşını sizce kim kazanır?

Fakat döngüde bir insan bulunduğu için inanılmaz derecede şanslı olduğumuz kıl payı durumlar da mevcut. 27 Ekim 1962'de, Küba Füze Krizi sırasında, on bir ABD Donanma destroyeri ve uçak gemisi USS *Randolph*, Sovyet denizaltısı B-59'u Küba yakınında, ABD "karantina" alanı dışındaki uluslararası sularda köşeye sıkıştırmıştı. Bilmedikleri şey ise denizaltının bataryaları bitmek üzere ve havalandırmanın durmuş olması sebebiyle güvertedeki sıcaklığın 45 °C'yi (113 °F) geçmiş olduğuydu. Karbondioksit zehirlenmesinin kıyısında, pek çok mürettebat bayılmıştı. Ekip Moskova'yla günlerdir bağlantı kurmamıştı ve Üçüncü Dünya Savaşı'nın başlayıp başlamadığını da bilmiyorlardı. Sonrasında Amerikalılar küçük su altı bombaları bırakmaya başladılar. Mürettebatın bundan haberi yoktu ama Moskova'ya bunun denizaltıyı yüzeye çıkıp gitmeye zorlamak için yapıldığını söylemişlerdi.

"Bunun son olduğunu düşündük," diye hatırlıyor mürettebattan V. P. Orlov. "Birinin sürekli çekiçle vurduğu metal bir varildeymişiz gibiydi." Amerikalıların bilmediği bir başka şey de B-59 ekibinin Moskova'dan onay almadan fırlatma izinleri olan nükleer bir torpidoya sahip olduğuydu. Gerçekten de Yüzbaşı Savitski nükleer torpidoyu fırlatmaya karar vermişti. Valentin Grigoryeviç, torpido subayı, şöyle bağırmıştı: "Öleceğiz ama hepsini de batıracağız – donanmamızı yüzüstü bırakmayacağız!" Neyse ki fırlatma kararının güvertedeki üç subay tarafından da onaylanması gerekiyordu ve biri, Vasili Arkhipov, hayır dedi. Kararı Üçüncü Dünya Savaşı'nı engellemiş ve modern tarihte insanlığa yapılmış en değerli katkı olmuş olmasına rağmen Arkhipov'u pek az kişinin duymuş olması üzücüdür.[37] B-59 bir otonom YZ kontrollü denizaltı olsaydı ve döngüde hiçbir insan bulunmasaydı ne olabilirdi diye düşünmek de ayrıca üzücüdür.

Yirmi yıl sonra, 9 Eylül 1983'te, süper güçler arasında yine sinirler gergindi: Sovyetler Birliği, ABD Başkanı Ronald Reagan tarafından "şeytani imparatorluk" olarak adlandırılmıştı ve bir önceki hafta, kendi hava sahasına girmiş bir Kore Havayolları yolcu uçağını düşürerek içinde bir ABD kongre üyesinin de bulunduğu iki yüz altmış dokuz kişiyi öldürmüştü. Şimdi de otomatik bir Sovyet erken uyarı sistemi ABD'nin Sovyetler Birliği'ne beş adet yerden fırlatılmış nükleer füze gönderdiğini söylüyordu. Bunun yanlış alarm olup olmadığına karar vermek için Subay Stanislav Petrov'un birkaç dakikası vardı. Uydu doğru çalışıyor gibi görünüyordu, bu yüzden de protokolü takip etmek yaklaşan bir nükleer saldırıyı rapor etmesini gerektiriyordu. Bunun yerine içgüdülerine güvendi ve ABD'nin yalnızca beş füzeyle saldırmasının mantıklı olmayacağına karar verdi ve doğru olup olmadığını bilmeden üstlerine bunun yanlış alarm olduğunu bildirdi. Sonradan uydunun, güneşin bulutlar üzerindeki yansımalarını

roket motorlarının alevleri sandığı ortaya çıktı.[38] Eğer Petrov yerine protokolü takip eden bir YZ sistemi olsaydı neler olurdu merak ediyorum.

Geleceğin Silahlanma Yarışı mı?

Şu ana kadar tahmin etmiş olacağınız üzere, kişisel olarak otonom silah sistemleri hakkında ciddi endişelerim var. Ancak size asıl endişemi anlatmaya başlamadım bile: YZ silahlarında silahlanma yarışının uç noktası. Temmuz 2015'te, Future of Life Institute'teki arkadaşlarımdan da aldığım faydalı geri bildirimlerle birlikte, Stuart Russell'la yazdığımız bir açık mektupta bu endişelerimi dile getirdim:[39]

OTONOM SİLAHLAR:

YZ & Robotik Araştırmacılarından Bir Açık Mektup

Otonom silahlar insan müdahalesi olmaksızın hedefleri seçer ve saldırır. Bunlar önceden tanımlanmış bazı kriterleri karşılayan, insanları arayıp yok edebilecek silahlı dört pervaneli helikopterleri içerir ancak insanların tüm hedef kararlarını verdiği cruise füzeleri ya da uzaktan kontrol edilen droneları içermez. Yapay zekâ (YZ) teknolojisi böyle sistemlerin kullanılmasının on yıllar değil yıllar içerisinde pratik hatta hukuki olarak uygun hale geleceği bir noktaya ulaştı ve sonuçları ağır olabilir: Otonom silahlar barut ve nükleer silahlardan sonra savaş alanında üçüncü devrim olarak anılıyor.

Otonom silahları savunan ya da eleştiren pek çok sav ileri sürüldü, örneğin insan askerleri makinelerle değiştirmenin zayiatı azaltmak açısından iyi olduğu ancak savaş çıkarma eşiğini

düşürdüğü için de kötü olduğu söylendi. Bugün insanlık için asıl soru küresel bir YZ silahlanma yarışı başlatmak mı yoksa bunun başlamasını engellemek mi olmalıdır. Eğer herhangi bir ana askerî güç YZ silah gelişiminde ilerlemeye başlarsa, küresel silahlanma yarışı kaçınılmaz olur ve bu teknolojik ilerlemenin uç noktası da açıktır: Otonom silahlar geleceğin kalaşnikofları olur. Nükleer silahların aksine, maliyeti yüksek ya da elde etmesi zor hammaddelere ihtiyaç duymazlar, bu yüzden de hazır ve nazır, tüm önemli askerî güçler için kitlesel üretime geçebilecek kadar ucuz olurlar. Karaborsaya düşmesi ve teröristlerin, kitlelerini daha iyi kontrol etmek isteyen diktatörlerin, etnik temizlik yapmak isteyen savaş baronlarının eline geçmesi işten bile değildir. Otonom silahlar suikastlar, ulusların dengesini bozmak, kitleleri baskı altına almak ve belli bir etnik grubu seçici biçimde öldürmek için idealdir. Bu yüzden de askerî bir YZ silahlanma yarışının insanlık için faydalı olmayacağına inanıyoruz. YZ'nin savaş alanlarını insanlar, özellikle de siviller için daha güvenli hale getirebileceği, insanları öldürmek için yeni araçlar oluşturmayabileceği pek çok başka yol mevcuttur.

Tıpkı pek çok kimyager ve biyoloğun kimyasal ya da biyolojik silah yapmaya istekli olmaması gibi, çoğu YZ araştırmacısı da YZ silahları yapmak konusunda istekli değildir ve başkalarının bunu yaparak kendi alanlarını kirleterek, gelecekte topluma sağlayacağı faydaları engelleyecek bir YZ karşıtı toplumsal hareketin ortaya çıkmasını da istememektedir. Gerçekten de kimyagerler ve biyologlar kimyasal ve biyolojik silahları başarılı biçimde engelleyen uluslararası anlaşmaları geniş çapta desteklerken, çoğu fizikçi de uzay merkezli nükleer silahları ve kör edici lazer silahları yasaklayan anlaşmaları desteklemiştir.

Endişelerimizin ağaç sevicilerin pasifist lafları diye görmezden gelinmesini zorlaştırmak için, mektubumuzun olabildiğince çok YZ araştırmacısı ve robotikçi tarafından imzalanmasını sağlamaya çalıştım. Robotik Silahlanma Kontrolü İçin Uluslararası Kampanya, katil robotların yasaklanması için çağrıda bulunan yüzlerce imzacıyı bir araya getirmişti ve ben daha iyisini yapabileceğimizi düşünüyordum. Profesyonel kurumların devasa üye e-posta listelerini politik olarak yorumlanabilecek bir amaç için paylaşmak istemeyeceklerini biliyordum. Ben de çevrimiçi belgelerden araştırmacı isimleri ve kurumlarının bir listesini bir araya getirdim ve e-posta adreslerini bulma görevi için Amazon kitle kaynak platformu MTurk'e ilan verdim. Çoğu araştırmacının e-posta adresleri üniversite web sitelerinde vardı ve yirmi dört saat ile elli dört dolar sonra, Yapay Zekânın İlerlemesi Derneği (AAAI) üyeleri olarak seçilebilecek kadar başarılı olan YZ araştırmacılarından yüzlercesinin e-postalarından oluşan bir listenin sahibiydim. Bunlardan biri, Britanyalı-Avustralyalı YZ profesörü Toby Walsh listesindeki herkese e-posta göndermeyi nazikçe kabul etti ve kampanyamızı ilerletmekte büyük katkı sundu. Dünyanın dört bir yanından MTurk çalışanları, Toby için devamlı ek e-posta listeleri hazırladılar ve kısa sürede üç binin üzerinde YZ ve robotik araştırmacısı açık mektubumuzu imzalamıştı. Bunların içinde AAAI'nın altı eski başkanı ve Google, Facebook, Microsoft ve Tesla'dan YZ endüstrinin liderleri de vardı. FLI gönüllü ordusu yorulmak bilmeden imzacı listelerini doğruluyorlar, Bill Clinton ve Sarah Connor gibi yanıltıcı isimleri ayırıyorlardı. Stephen Hawking de dâhil olmak üzere on yedi binin üzerinde kişi imzaladı ve Toby, Uluslararası Ortak Yapay Zekâ Konferansı'nda bununla ilgili bir basın açıklaması organize ettiğinde, tüm dünyada haber olduk.

Çünkü biyologlar ve kimyagerler zamanında duruş sergiledikleri için, alanları şu anda faydalı ilaçlar ve materyaller üretmekle biliniyor, biyolojik ve kimyasal silahlar değil. YZ ve robotik çevreleri de sözlerini söylediler: Mektubu imzalayanlar da kendi alanlarının daha iyi bir gelecek oluşturmakla tanınmasını istiyor, insanları öldürmenin yeni yollarını yaratmakla değil. Fakat YZ'nin gelecekteki ana kullanım alanı sivil mi olacak askerî mi? Bu bölümde ilkine daha fazla sayfa ayırmış olmamıza rağmen, ikincisine daha fazla para harcayacağız gibi görünüyor – özellikle de askerî YZ silahlanma yarışı başlarsa. Sivil YZ yatırım taahhütleri 2016'da milyar doları aşmıştı ancak bu Pentagon'un 2017 mali bütçe talebinde YZ bağlantılı projeler için istediği on iki ila on beş milyar doların yanında pek bir şey değil. Üstelik, Çin ve Rusya da, Savunma Bakan Yardımcısı Robert Work'ün şu sözlerini not etmiş olmalıdır: "Rakiplerimizin siyah perdenin ardında ne olduğunu merak etmelerini istiyorum."[40]

Uluslararası Bir Anlaşma Olmalı mı?

Şu an katil robot yasağını tartışmaya yönelik ciddi bir uluslararası baskı olsa da ne olacağı hâlâ daha açık değil ve olacaksa da ne olması *gerektiği* hakkında canlı bir tartışma sürüyor. Pek çok önemli paydaş dünya güçlerinin OSS araştırması ve kullanımını yönetmek için bir tür uluslararası regülasyona başvurması gerektiği konusunda hemfikir olsa da tam olarak neyin yasaklanması gerektiği ve yasağın nasıl uygulanacağı üzerine pek az uzlaşma var. Mesela, yalnızca ölümcül otonom silahlar mı yasaklanmalı yoksa insanları ciddi anlamda yaralayan, mesela kör eden, silahlar da yasaklanmalı mı? Geliştirme, üretim ya da sahip olmayı mı yasaklamalıyız? Yasaklar tüm otonom silah sistemlerine uygulanmalı mı yoksa mektubumuzda belirttiğimiz gibi, yalnızca saldırı

amaçlı olanlara uygulanmalı ve otonom uçaksavar silahları ve savunma füzeleri gibi savunma sistemlerine izin verilmeli mi? İkinci durumda OSS düşman bölgesine hareket ettirilmesi kolay olduğu zamanlarda bile savunma olarak mı değerlendirilmeli? Ve bir otonom silahın çoğu bileşeninin aynı zamanda sivil işlevi de varken bir yasayı nasıl uygulayabilirsiniz? Mesela Amazon paketleri taşıyan bir drone ile bomba atan arasında çok da bir fark yoktur.

Bazı kişiler etkili bir OSS anlaşması tasarlamanın umutsuzca zor olduğunu ve denemeye bile gerek olmadığını söylüyorlar. Öte yandan, John F. Kennedy Ay görevlerini açıklarken başarının insanlığın geleceği için büyük fayda sağlayacağı durumlarda zor şeyleri denemenin değeceği yönündeki açıklaması vardır. Dahası, çoğu uzman biyolojik ve kimyasal silahlara uygulanan yasakların, çok fazla hile olduğu için uygulanmaları zor bile olsa değerli olduklarını söylüyor çünkü yasaklar, kullanımlarını sınırlayan ciddi yaftalamalara yol açmıştı.

2016'da bir yemekli davette Henry Kissinger'la tanıştım ve ona biyolojik silahların yasaklanmasındaki rolünü sorma şansına eriştim. ABD ulusal güvenlik danışmanıyken Başkan Nixon'ı yasağın ABD ulusal güvenliği için iyi olabileceğine ikna ettiğini anlattı. Doksan iki yaşındayken bile hafızasının bu kadar keskin olmasından etkilenmiş ve olayın içyüzünü duyduğum için mutlu olmuştum. ABD konvansiyonel ve nükleer kuvvetleri sayesinde süper güç statüsünü koruduğu için, sonucu belli olmayan bir dünya çapında biyolojik silah yarışında kazanacağından çok kaybedeceği şey vardır. Diğer bir deyişle, zaten zirvedeysen, "bozuk değilse, tamir etme" düsturunu takip etmek anlamlıdır. Stuart Russell yemek sonrası konuşmamıza katıldı ve aynı savın ölümcül otonom silahlar için de yapılabileceğini tartıştık: Silahlanma yarışından en çok kazanç ihtimali olanlar süper güçler değil küçük haydut

devletler ve geliştirildikleri anda karaborsadan silahlara ulaşabilecek teröristler gibi devlet dışı aktörlerdir.

Kitlesel üretime geçildiğinde, küçük YZ destekli katil dronelar bir akıllı telefondan biraz daha pahalı olur. İster bir politikacıya suikast gerçekleştirmek isteyen bir terörist ister eski kız arkadaşından intikam almak isteyen bir sevgili olsun, yapması gereken tek şey hedefin fotoğrafını ve adresini katil dronea yüklemek olacaktır: Hedefe uçar, tanımlar, o kişiyi ortadan kaldırır ve kimsenin kimin sorumlu olduğunu bilmemesi için kendini yok eder. Alternatif olarak, etnik temizliğe meyilli olanlar için, yalnızca deri rengi ya da etnisiteye göre insanları öldürebilecek şekilde programlanabilirler. Stuart bu silahların akıllandıkça, her bir ölüm için daha az materyal, ateş gücü ve para gerektireceğini ileri sürüyor. Mesela, arı boyutunda droneların insanları gözünden vurarak minimum patlayıcı kuvvetiyle ucuza öldürebileceğinden korkuyor. Göz yeterince yumuşak olduğu için beyne kadar devam edebilecek harekete de izin verebilir. Ya da metal pençelerle kafaya tutunabilirler ve kafatasını delebilirler. Tek bir kamyonun arkasından bir milyon tane böyle katil drone salınırsa, elimizde yepyeni bir tür korkunç bir kitle imha silahı var demektir: Seçici bir biçimde yalnızca önceden belirlenmiş bir insan kategorisini öldürebilen bir silah, hem de başka kimseye ve hiçbir şeye zarar vermeden.

Sık rastlanan bir karşı sav da böyle endişelerden kurtulmak için katil robotları etik yapmaktır, yani öyle robotlar yaparız ki yalnızca düşman askerleri öldürürler. Fakat bir yasak uygulamak konusunda endişeleniyorsak, düşman otonom silahlarının %100 etik olmasını sağlamanın hiç üretilmemesini sağlamaktan daha kolay olduğundan nasıl emin olabiliriz? Ayrıca, aynı anda haydut devletlerin, diktatörlerin ve terörist grupların savaş kurallarını çok iyi uyguladıkları için bu kuralları ihlal etmeyecek şekilde

robot kullanmayı asla seçmeyeceklerini iddia ederken, medeni ulusların iyi eğitimli askerlerinin savaş kurallarını uygulamakta kötü olduğunu söyleyip robotların daha iyi yapabileceğini ileri sürmek mümkün müdür?

Siber Savaş

YZ'nin bir diğer ilgi çekici askerî yanı da hiçbir silah yapmadan bile düşmanınıza saldırabilecek olmanızdır, yani siber savaş. Geleceğin neler getirebileceğine dair küçük bir giriş olarak, genelde ABD ve İsrail hükümetlerine atfedilen Stuxnet solucanı İran'ın nükleer zenginleştirme programındaki hızlı dönen santrifüjleri etkilemiş ve kendilerini parçalamalarına yol açmıştı. Toplum ne kadar otomatikleşirse YZ'ye saldırmak o kadar güçlü etki yaratır ve siber savaş da o kadar yıkıcı olur. Eğer düşmanınızın sürücüsüz arabalarına, otopilottaki uçaklarına, nükleer reaktörlerine, endüstriyel robotlarına, iletişim sistemlerine, finansal sistemlerine ve enerji ağlarına sızarsanız, ekonomisini çökertip savunmasını düşürebilirsiniz. Eğer silah sistemlerine sızarsanız, daha da iyi.

Bu bölüme YZ'nin insanlığa fayda sağlayacak yakın dönem imkânlarının ne kadar baş döndürücü olduğunu inceleyerek başladık, tabii eğer sağlam ve hacklenemez inşa edebilirsek. YZ'nin kendisi YZ sistemlerini daha sağlam yaparak siber savaş savunmasına yardımcı olabilecek olsa da saldırıya da yardımcı olabilir. YZ gelişimi için savunmanın başarılı olmasını garanti altına almak en hayati kısa dönem hedeflerinden biridir. Aksi halde inşa ettiğimiz tüm harika teknoloji aleyhimize çevrilebilir!

İşler ve Ücretler

Bu bölümde şu ana kadar YZ'nin bütçeye uygun fiyatlarda dönüştürücü ve yeni ürünlerle hizmetler sunarak *tüketiciler* olarak

bize nasıl etki edeceğini gördük. Peki, iş pazarını dönüştürerek *çalışanlar* olarak bize nasıl etki edecek? İnsanları gelirlerinden ve amaçlarından etmeden otomasyon yoluyla refahımızı artırmanın bir yolunu bulabilirsek, isteyen herkes için eğlence ve önceden görülmemiş servet ile fantastik bir gelecek yaratma potansiyeline sahip oluruz. Bu konu üzerine çok az insan, MIT'deki çalışma arkadaşlarımdan biri olan ekonomist Erik Brynjolfsson kadar uzun ve etraflıca düşünmüştür. Hep bakımlı ve kusursuz giyimli olsa bile, İzlanda kökenli olduğu için, işletme okulumuza ayak uydurmak için söz dinlemez kızıl Viking sakalını yeni kesmiş gibi gelir. Çılgın fikirlerini tıraş etmediği ise kesindir ve iyimser iş pazarı vizyonuna "Dijital Atina" adını verir. Eski çağlarda Atinalı vatandaşların demokrasi, sanat ve sporu aynı anda deneyimleyebildikleri eğlence dolu bir yaşam sürebilmiş olmalarının sebebi aslında işin büyük kısmını kölelerin yapmasıydı. Peki, herkesin mutlu olabileceği bir dijital ütopya yapmak için YZ destekli robotları kölelerin yerine koymaya ne deriz? Erik'in YZ merkezli ekonomisi stresi, angaryayı kaldırmak ve bugün istediğimiz her şeyi bollukla üretmekle kalmaz, ayrıca bugünün tüketicilerinin henüz istediklerinin farkında olmadıkları bol miktarda yeni ürün ve hizmeti de üretebilir.

Teknoloji ve Eşitsizlik

Eğer herkesin saatlik maaşı her yıl artmaya devam ederse, böylece de devamlı olarak yaşam standartlarını artırarak daha fazla boş zaman isteyenlerin daha az çalışabileceği bir durum yaratılırsa Erik'in Dijital Atina'sına bugün olduğum yerden varabiliriz. Görsel 3.5 İkinci Dünya Savaşı'ndan 1970'lerin ortasına kadar ABD'de tam olarak bunun gerçekleştiğini gösteriyor: Gelir eşitsizliği olsa da pastanın toplam büyüklüğü herkesin daha büyük bir dilim

alabileceği şekilde artmıştı. Ancak sonra, ilk olarak Erik'in gösterdiği gibi bir şeyler değişti: Görsel 3.5, ekonominin büyümesine ve ortalama geliri yükseltmesine rağmen, son kırk yıldaki kazanımların çoğunun en zengine, çoğunlukla da en üstteki %1'e gittiğini ve en fakir %99'un gelirlerinin durağanlaştığını gösteriyor. Eşitsizlikte artan büyüme, gelire değil zenginliğe baktığımızda çok daha belirgin oluyor. ABD evlerinin en aşağıdaki %90'ı için, ortalama net değer 2012'de yaklaşık seksen beş bin dolarken –yani yirmi beş yıl öncesi ile aynıyken– en üstteki %1 bu sürede enflasyondan arındırılmış zenginliklerini iki katından fazlasına, on dört milyon dolara çıkarmıştır.[41] Uluslararası arenada farklar daha da yüksektir. 2013'te dünya nüfusunun alt yarısının (yani 3.6 milyardan fazla insanın) tüm zenginliği dünyanın en zengin sekiz insanının zenginliğine eşitti.[42] Üstteki zenginliğe olduğu kadar aşağıdaki fakirlik ve hassasiyete de vurgu yapan bir istatistik bu. 2015'teki Porto Riko Konferansı'mızda, Erik bir araya gelmiş YZ araştırmacılarına YZ ve otomasyondaki gelişmelerin ekonomik pastayı büyüteceğini ancak herkesin, hatta çoğu kimsenin bile, bundan fayda sağlayacağına dair hiçbir ekonomik yasa olmadığını söyledi.



Görsel 3.5: Ekonominin ortalama geliri son yüzyılda nasıl artırdığını ve bu gelirin yüzde kaçının farklı gruplara gittiğini göstermektedir. 1970'lerden önce, zenginler ve fakirler birbirleriyle uyumlu biçimde daha iyi kazanmaktaydılar, daha sonra kazanımların çoğu en üst %1'e giderken en alt

%90 ortalama sıfıra yakın ek kazanç elde etti.[43] Rakamlar 2017 dolarına göre enflasyon hesaba katılarak düzenlenmiştir.

Ekonomistler arasında eşitsizliğin arttığına dair ciddi bir fikir birliği olsa da bu trendin neden devam ettiğine ve devam edip etmeyeceğine dair enteresan bir tartışma da var. Politik spektrumun solunda kalan tartışmacılar ana sebebin küreselleşme ve/veya zenginler için vergi indirimleri gibi ekonomik politikalar olduğunu savunuyorlar. Ancak Erik Brynjolfsson ve MIT'deki çalışma arkadaşı Andrew McAfee ana sebebin başka bir şey olduğunu düşünüyorlar: teknoloji.[44] Özellikle de dijital teknolojinin eşitsizliği üç farklı şekilde yarattığını öne sürüyorlar.

İlk olarak, eski işleri daha fazla beceri gerektiren işlerle değiştirerek, teknoloji eğitimli olanları ödüllendirmiştir: 1970'lerin ortalarından beri, üniversite derecesi olanlar için maaşlar %25 artarken, ortalama bir lise terk öğrencininki %30 kesintiye uğramıştır.[45]

İkincisi, 2000 yılından beri, daha da büyük bir kurumsal gelir paydası o kurumlarda çalışanlara değil kurumların sahiplerine gitmeye başladı ve otomasyon sürdüğü sürece de makinelere sahip olanların pastadan giderek daha büyük bir pay alacaklarını söyleyebiliriz. Sermayenin emek üzerindeki bu üstünlüğü, teknoloji vizyoneri Nicholas Negroponte tarafından atomlar değil, hareket eden parçalar olarak tanımlanan, büyüyen dijital ekonomi için özellikle önemli olabilir. Kitaplardan filmlere ve vergi hazırlama araçlarına her şeyin dijital olduğu bir zamanda, ek kopyalar hiçbir ekstra çalışana ihtiyaç duyulmaksızın neredeyse sıfır maliyetle tüm dünyada satılabilir. Bu da çoğu kârın çalışanlardan çok yatırımcılara gitmesini sağlar. Aynı şekilde bu, 1990 yılında Detroit'in "büyük üçlüsünün" (GM, Ford ve Chrysler) toplam gelirlerinin 2014 yılında Silikon Vadisi'nin "büyük üçlüsünün" (Google, Apple,

Facebook) toplam gelirine, ikinci grubun diğerine göre dokuz kat daha az çalışanı olmasına rağmen nasıl eşit olabileceğini ve borsada nasıl otuz kat daha değerli olabileceğini de açıklamaya yardımcı olur.[46]

Üçüncü olarak, Erik ve arkadaşları süperstarların dijital ekonomide herkesten daha şanslı olduğunu ileri sürüyor. *Harry Potter* yazarı J. K. Rowling milyarderler kulübüne adım atan ilk yazar oldu ve Shakespeare'den daha çok para kazanıyor çünkü hikâyeleri yazı, film ve oyun formunda milyarlarca insana çok düşük maliyetle iletilebiliyor. Aynı şekilde Scott Cook, TurboTax vergi hazırlama yazılımıyla milyar kazandı çünkü insan vergi hazırlayıcılarının aksine bu yazılım indirilebiliyordu. Çoğu insan en iyi onuncu vergi hazırlama yazılımına çok az ya da hiç para ödememeye yatkın olacağı için, pazarda ancak birkaç süperstara yer vardır. Bu da dünyadaki tüm ebeveynler çocuklarına bir sonraki J. K. Rowling, Gisele Bündchen, Matt Damon, Cristiano Ronaldo, Oprah Winfrey ya da Elon Musk olmalarını söyleseler bile, çocukların neredeyse hiçbirinin bunu mantıklı bir kariyer stratejisi olarak görmeyeceği anlamına gelir.

Çocuklar İçin Kariyer Tavsiyesi

Peki, çocuklarımıza nasıl bir kariyer tavsiyesi *vermeliyiz*? Ben kendi çocuğumu makinelerin şu an yapamadığı ve böylece yakın gelecekte otomatikleştirilmeyecek mesleklere yönelmeye teşvik ediyorum. Çeşitli işlerin makinelerce ne zaman ele geçirileceğine dair tahminler insanların bir kariyer seçmeden önce sorabileceği birkaç faydalı soru ortaya koyuyor.[47] Mesela:

- İnsanlarla etkileşime geçmeyi ve sosyal zekâ kullanımını gerektiriyor mu?

- Yaratıcılığı ve zeki çözümler getirmeyi içeriyor mu?
- Kestirilemez bir çevrede çalışmayı gerektiriyor mu?

Bu tip sorulardan ne kadar çoğuna evet diyebiliyorsanız, kariyer seçiminiz o kadar iyi olur. Yani görece güvenli olanlar öğretmenlik, hemşirelik, doktorluk, diş hekimliği, bilim insanlığı, girişimcilik, programcılık, mühendislik, avukatlık, sosyal hizmetler, din adamlığı, sanatçılık, kuaförlük ya da masaj terapistliği gibi mesleklerdir.

Öte yandan yüksek oranda tekrar ve yapısallaştırılmış eylemleri tahmin edilebilir bir çevrede yapmayı içeren mesleklerin çok uzun süre dayanmayacağı kesindir. Bilgisayarlar ve endüstriyel robotlar böyle basit işleri uzun süre önce ele geçirmişti ve gelişen teknoloji telepazarlamacılardan depo çalışanları, kasiyerler, tren operatörleri, fırıncılar ve yan aşçılar gibi birçok işi de ortadan kaldırmak üzere.[48] Kamyon, otobüs, taksi ve Uber/Lyft araçlarının şoförleri de onları takip edecek gibi görünüyor. Soyu tükenecekler listesinde olmasalar bile pek çok görevi otomatikleşmiş olduğu için daha az insana gerek duyacak bazı meslekler (mesela, avukat asistanları, kredi analistleri, borç memurları, muhasebeciler ve vergi memurları gibi) de vardır.

Otomasyondan uzak durmak kariyerdeki tek zorluk değildir. Bu küresel dijital çağda, profesyonel bir yazar, sinemacı, aktör, sporcu ya da moda tasarımcısı olmayı hedeflemek başka bir sebepten ötürü risklidir: Bu mesleklerdeki insanlar kısa sürede makinelerle ciddi bir rekabete girmeyecek olsalar da yukarıda bahsettiğimiz süperstar teorisinden ötürü dünyadaki diğer insanlarla artan bir rekabete girecekler ve pek azı başarı kazanacaktır.

Çoğu vakada, tüm bir alan konusunda kariyer tavsiyesi vermek miyop ve acımasızca olacaktır: Tamamen ortadan kalkmayacak

ama bazı görevlerinin otomatikleşeceği pek çok iş vardır. Mesela, tıp alanına girerseniz, medikal görüntüleri analiz edip IBM'in Watson'ı tarafından işinden edilen radyolog olmayın, radyoloji analizini isteyen, sonuçları hastayla tartışan ve tedavi planına karar veren kişi olun. Finans alanına girerseniz, veriye algoritmaları uygulayan ve yazılım tarafından işinden edilen kişi olmayın, sayısal analiz sonuçlarını stratejik yatırım kararları vermekte kullanan fon yöneticisi olun. Eğer hukukçu olacaksanız, bulgulama süreci için binlerce dokümanı inceleyen avukat asistanı olmayın, müvekkiline fikir veren, davasını mahkemede sunan avukat olun.

YZ çağında bireylerin iş pazarında başarılarını nasıl maksimize edebileceklerine baktık. Peki, hükümetler iş gücünün başarıya ulaşması için neler yapabilir? Mesela, hangi eğitim sistemi insanları YZ'nin hızla ilerlediği bir iş pazarına daha iyi hazırlar? On ya da yirmi yıllık eğitimi müteakip kırk yıllık özelleşmiş çalışmadan oluşan mevcut modelimiz mi? Yoksa insanların bir süre çalıştığı, sonra bir yıllığına okula döndüğü ve tekrar birkaç yıl çalıştığı bir sisteme mi geçmeliyiz?[49] Ya da sürekli eğitim (muhtemelen çevrimiçi) herhangi bir işin standart parçası mı olmalı?

Yeni iyi işler yaratmakta en faydalı olacak ekonomik politikalar nelerdir? Andrew McAfee araştırma, eğitim ve altyapıya ciddi oranda yatırım yapmaktan göçü kolaylaştırma ve girişimciliği teşvik etmeye dek pek çok yardımcı olabilecek politika olduğunu söylüyor. "Ekonomi 101'in gösterdikleri çok açık ama takip edilmiyor," diye düşünüyor, en azından ABD'de.[50]

Eninde Sonunda İnsanlar İşe Alınamaz Hale mi Gelecek?

Eğer YZ gelişimini sürdürür, daha da fazla işi otomatikleştirirse, ne olur? Çoğu insan otomatik işlerin daha da iyi olan yenileriyle değiştirileceğini söyleyerek iş konusuna iyimser yaklaşıyorlar.

Sonuçta daha önce de benzeri olmuştu, Ludditeların Sanayi Devrimi'nde işsizlikten endişe etmelerinden beridir yani.

Diğerleriyse iş konusunda karamsarlar ve bu sefer her şeyin daha farklı olacağını, çok daha fazla insanın yalnızca işsiz kalmayacağını, işe alınamaz hale de geleceğini söylüyorlar.[51] Karamsarlar serbest piyasanın maaşı arz talep dengesine göre belirlediğini, artan ucuz makine emeği arzının insan maaşlarını yaşama maliyetinin çok altına düşüreceğini ileri sürüyorlar. Bir iş için piyasa maaşı bu işi en ucuza yapanın saatlik maliyeti olduğu için maaşlar, tarih boyunca ne zaman belirli bir işi daha düşük gelirli bir ülkeye ya da daha ucuz bir makineye yaptırma fırsatı çıksa düşmüştür. Sanayi Devrimi esnasında, kaslarımızı makinelerle değiştirmeyi nasıl başaracağımızı öğrendik ve insanlar da daha çok zihinlerini kullandıkları daha yüksek maaşlı işlere geçtiler. Mavi yaka işlerin yerini beyaz yaka işler aldı. Şu anda zihinlerimizin yerini makinelerin nasıl alabileceğini çözmeye başlıyoruz. Eğer bunda da başarılı olursak, bize ne iş kalır ki?

Bazı iyimserler, fiziksel ve zihinsel işlerin ardından bir sonraki patlama *yaratıcı* işlerde olacak diyor ancak karamsarlar yaratıcılığın da başka bir zihinsel süreç olduğunu, bunun da YZ tarafından yapılabileceğini ileri sürüyorlar. Diğer iyimserler bir sonraki patlamanın yeni teknolojinin ortaya çıkaracağı henüz daha aklımıza bile gelmeyen işlerde olacağını umuyorlar. Sonuçta, Sanayi Devrimi zamanında kim torunlarının bir gün web tasarımcısı ya da Uber sürücüsü olacağını düşünebilirdi ki? Fakat karamsarlar bunun hiçbir ampirik veriye dayanmayan hüsnükuruntu olduğunu söyleyerek fikre karşı çıkıyor. Söyledikleri şey aynı savı yüz yıl önce bilgisayar devriminden önce de yapabileceğimiz ve bugünün mesleklerinin çoğunun yeni ve daha önceden var olmayan hayal edilmemiş teknoloji işleri olacağını tahmin edebileceğimizdir. Böyle bir tahmin büyük bir yanlış olurdu elbette, Görsel 3.6'da da görebileceğiniz üzere: Bugünün işlerinin büyük

bir kısmı yüz yıl önce de mevcuttu ve sundukları iş olanaklarına göre sıraladığımızda, yeni bir iş bulana dek yirmi birinci sıraya kadar inmemiz gerekiyor: Bu da yazılım geliştiriciliktir, ABD iş piyasasının yüzde birinden azını kaplar.

İnsan zekâsının manzarasını makineler için çeşitli görevleri yapmanın ne kadar zor olduğunu gösteren yükseltilerle ve makinelerin şu an yapabildiği şeyleri yükselen deniz seviyesiyle temsil eden, Bölüm 2'deki Görsel 2.2'yi hatırlarsak neler olduğuna dair daha iyi bir anlayışa sahip olabiliriz. İş piyasasındaki ana trend yepyeni işlere ilerliyor oluşumuz değildir. Aksine, yükselen teknoloji dalgasının sular altında bırakmadığı Görsel 2.2'deki o kara parçalarına sıkışıyoruz! Görsel 3.6 bunun tekil bir ada değil, halen daha makinelerin insanlar kadar ucuza yapamadığı tüm o değerli şeylere karşılık gelen adacıklar ve mercanlarla dolu karmaşık bir takımadada olduğunu gösteriyor. Bu yalnızca yazılım geliştirme gibi yüksek teknoloji meslekleri değil, masaj terapisinden aktörlüğe kadar üstün maharet ve sosyal becerimizi güçlendiren düşük teknoloji bir dizi işi de kapsıyor. Belki de YZ entelektüel görevlerde bizi o kadar hızlı saf dışı bırakacak ki kalan tek işler düşük teknoloji kategorisinde olacak. Arkadaşlarımdan biri bana en sona kalan mesleğin ilk meslek olacağını söyledi: seks işçiliği. Ancak Japon bir robotikçiye bundan bahsedince o hemen karşı çıktı: "Hayır, robotlar da bu işlerde çok iyidir!"

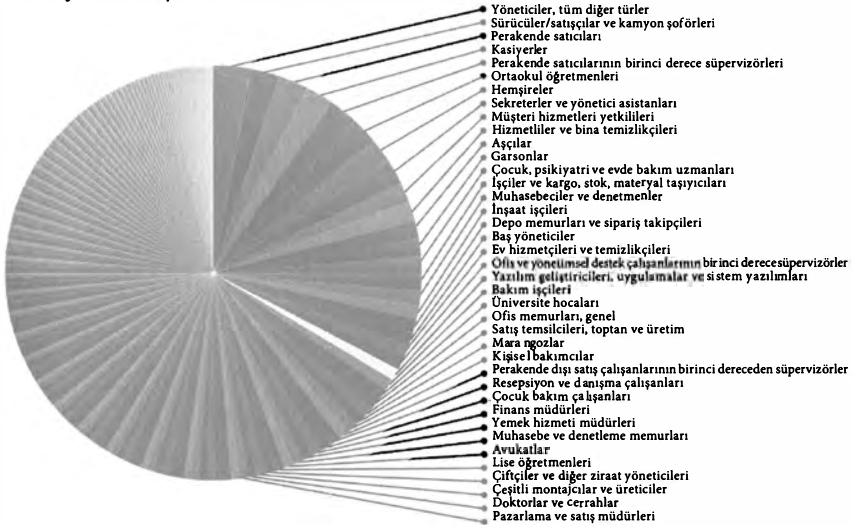
İş karamsarları sonun kaçınılmaz olduğunu düşünüyorlar: Tüm takımada sular altında kalacak ve insanların makinelerden daha ucuza yapabildiği hiçbir iş kalmayacak. 2007 yılında çıkan kitabı *Farewell to Alms*'ta İskoç-Amerikan ekonomist Gregory Clark atlara bakarak gelecekteki iş imkânlarımız hakkında birkaç şey öğrenebileceğimize dikkat çekmişti. 1900 yılında erken dönem otomobillerden birine bakan ve geleceklerini değerlendiren iki atı düşünün.

"Teknolojik işsizlik konusunda endişelerim var."

"İhihih, Ludditelik yapma: Atalarımız da aynı şeyi buhar motorları sanayi işlerimizi, trenler de at arabasına koşulma işlerimizi aldığında aynı şeyi demişti. Ama şu an eskisinden daha çok işimiz var, hem de daha iyi işler. Şehirde küçük bir aracı çekmeyi tüm gün salak bir maden kuyusu pompasını çalıştırmak için daireler çizmeye tercih ederim."

"Peki ya bu içten yanmalı motor işi iyice yerleşirse?"

"Hiç hayal etmediğimiz yeni at işleri çıkacaktır eminim. Daha önce de hep aynısı oldu, tekerleğin ya da sabanın keşfindeki gibi."



Görsel 3.6: Daire grafiği 2015 yılında işi olan 149 milyon Amerikalının işlerini, popülerliğe göre sıralanmış şekilde ABD İş İstatistikleri Bürosu'ndan alınmış 535 iş kategorisiyle göstermektedir.[52] Bir milyondan fazla çalışanı olan tüm meslekler etiketlenmiştir. Yirmi bir numaraya kadar bilgisayar

teknolojisi tarafından oluşturulmuş hiç yeni iş bulunmamaktadır. Bu şekil Federico Pistono'nun bir analizinden esinlenmiştir.[53]

Maalesef bu henüz hayal edilmemiş yeni at işleri hiçbir zaman ortaya çıkmadı. Artık pek ihtiyaç duyulmayan atlar kesime gönderildi ve yerlerine yenileri gelmedi. Bu da 1915'te yirmi altı milyon olan ABD at popülasyonunun 1960'ta üç milyona düşmesine sebebiyet verdi.[54] Mekanik kasların atları gereksiz yaptığı gibi, mekanik zihinler de aynı şeyi insanlara yapacak mı?

İnsanlara İş Olmadan Gelir Vermek

Peki, şimdi kim haklı: Otomatikleşmiş işlerin daha iyileriyle değiştirileceğini söyleyenler mi yoksa çoğu insanın işe alınamaz hale geleceğini söyleyenler mi? Eğer YZ gelişimi yavaşlamak bilmeden devam ederse, o zaman *iki* taraf da haklı olabilir: biri kısa vadede ve diğeri de uzun vadede. Ancak insanlar sık sık işlerin ortadan kalkmasını bir felaket senaryosu gibi görseler de kötü bir şey olmak zorunda değil! Ludditelar bazı işler konusunda takıntılıydılar, başka işlerin de aynı sosyal değeri sağlayabilme ihtimalini görmezden geldiler. Benzer şekilde, belki de bugünkü işler konusunda takıntılı olanlar da biraz dar görüşlü olabilirler: İş istiyoruz çünkü bu iş bize gelir ve amaç veriyor ama makineler tarafından sağlanan bol kaynağı düşününce, hem gelir hem de işler *olmaksızın* hayatta bir amaç bulmanın alternatif yolları var olabilir. At hikâyesindekine benzer bir şekilde sona erebilir, sonuçta atların hepsinin soyu tükenmemişti. Bunun yerine, 1960'tan bu yana atların sayısı üç katına çıktı çünkü bir tür sosyal refah sistemi içinde korunuyorlardı: Kendi faturalarını ödeyemeseler de insanlar atlara bakmaya karar verdiler, eğlence, spor ya da

dostluk için onları yanlarında tuttular. Biz de aynı şekilde muhtaç insan dostlarımızı gözetemez miyiz?

Gelir meselesiyle başlayalım: Artan ekonomi pastasının ufak bir kısmını yeniden dağıtmak bile herkesin daha iyi gelire sahip olmasını sağlamalıdır. Pek çok kimse bunu *yapabilmenin* ötesinde yapmak *zorunda* olduğumuzu söylüyor. Moşe Vardi'nin YZ destekli teknolojiyle yaşamları kurtarmanın ahlaki zorunluluğundan bahsettiği bir 2016 panelinde, zenginliğin paylaşılması da dâhil olmak üzere faydalı kullanımını savunmanın da ahlaki bir zorunluluk olduğunu ileri sürmüştüm. Aynı panelde bulunan Erik Brynjolfsson, "Bu yeni zenginlik nesliyle bile insanların yarısının daha kötü bir gelir kazanmasını engelleyemiyorsak bize yazıklar olsun," demişti.

Zenginliğin paylaşılması için pek çok farklı teklif var ve her birinin hem destekçileri hem de karşıtları mevcut. En basiti *vatandaşlık maaşı.* Bu sistemde herkes hiçbir koşul ve gereklilik olmaksızın aylık bir ödeme alır. Kanada, Finlandiya ve Hollanda gibi bazı ülkelerde küçük çaplı bir dizi deneme yapıldı ya da yapılması planlandı. Bu sistemin taraftarları vatandaşlık maaşının ihtiyaç sahiplerine sosyal güvenlik ödemeleri yapmak gibi alternatiflerden daha verimli olduğunu savunuyorlar çünkü kimin hak ettiğine karar vermek için harcanan yönetimsel güçlükleri ortadan kaldırıyor. İhtiyaç temelli ödemeler işten caydırıcı olmakla da eleştiriliyor ancak kimsenin çalışmadığı işsiz bir gelecekte bu eleştirinin bir hükmü kalmayacak.

Hükümetler vatandaşlarına yalnızca para vererek değil, yollar, köprüler, parklar, toplu taşıma, çocuk bakımı, eğitim, sağlık, huzurevleri ve internet erişimini ücretsiz ya da sübvanse edilmiş şekilde sağlayarak da yardımcı olabilir. Gerçekten de pek çok devlet zaten bu hizmetlerin büyük bir kısmını sunuyor. Vatandaşlık maaşının aksine, böylesi hükümet destekli hizmetler iki ayrı hedefi gerçekleştiriyor: İnsanların yaşam maliyetlerini

azaltıyor ve iş sağlıyor. Makinelerin insanları tüm işlerde geçtiği bir gelecekte bile, hükümetler bakıcılık işlerini robotlara vermek yerine insanlara çocuk ve yaşlı bakımı *gibi* işlerde çalışmaları için ödeme yapabilir.

İlginç bir biçimde, devlet müdahalesi olmasa bile teknolojik ilerleme pek çok değerli ürün ve hizmeti ücretsiz sağlama noktasına gelebilir. Mesela, insanlar eskiden ansiklopedilere, atlaslara, mektup göndermeye ve telefon aramalarına para öderdi ancak şimdi internet bağlantısı olan herkes tüm bunlara sıfır maliyetle erişebiliyor, hem de ücretsiz video konferans, fotoğraf paylaşımı, sosyal medya, çevrimiçi dersler ve sayısız diğer yeni hizmet de cabası. Bir kişi için hayli değerli olabilen pek çok başka şey, mesela antibiyotikler üzerine hayat kurtarıcı bir ders, olabildiğince ucuz hale geldi. Teknoloji sayesinde, pek çok fakir insan bile geçmişte dünyanın en zengin insanlarının erişimi olmayan şeylere sahip. Bazıları bunun düzgün bir yaşam sürmek için gereken gelirin düştüğü anlamına geldiğini düşünüyor.

Eğer makineler bir gün tüm mevcut ürün ve hizmetleri minimum maliyetle üretebilirlerse, o zaman herkesin daha iyi yaşaması için yeterince zenginlik olacaktır. Diğer bir deyişle, görece mütevazı vergiler bile devletlerin vatandaşlık maaşı ve ücretsiz hizmetler sunması için yeterli olur. Ancak zenginliğin *paylaşılabileceği* gerçeği her hâlükârda *paylaşılacağı* anlamına gelmiyor ve olması *gerekip* gerekmediğine dair bile bugün ciddi bir politik uyuşmazlık var. Yukarıda gördüğümüz gibi, ABD'deki mevcut trend ters yönde gibi görünüyor, halk kitleleri her yıl giderek daha da fakirleşiyor. Toplumun artan zenginliğinin nasıl paylaşılacağına dair siyasi kararlar herkesi etkiliyor, bu yüzden de ne tür bir gelecek ekonomisi inşa etmek gerektiğine dair tartışma

da yalnızca YZ araştırmacıları, robotikçiler ve ekonomistleri değil herkesi ilgilendiriyor.

Pek çok tartışmacı, gelir eşitsizliğini azaltmanın yalnızca YZ'nin domine ettiği bir gelecekte değil şu anda da iyi bir fikir olduğunu savunuyor. Ana sav ahlaki olsa da yüksek derece eşitliğin demokrasinin daha iyi çalışmasını sağlayacağına dair kanıt da mevcut. Daha büyük, iyi eğitimli bir orta sınıf olursa, seçmen kitlesini manipüle etmek daha zor olur ve bir grup insanın ya da şirketin hükümet üzerinde haksız etki yaratmasının önüne geçilir. Daha iyi demokrasi, karşılığında, daha az yozlaşmış, daha verimli, daha hızlı büyüyen ve nihayetinde herkesin faydalanabildiği daha iyi yönetilen bir ekonominin yolunu açar.

İşler Olmadan İnsanlara Amaç Vermek

İşler insanlara paradan fazlasını verebilir. Voltaire 1759'da, "İş üç büyük şeytanı uzakta tutar: sıkıntı, ahlaksızlık ve ihtiyaç," demiştir. Öte yandan, insanlara gelir sağlamak esenliklerini garanti altına almaya yetmez. Roma imparatorları astlarını mutlu etmek için onlara hem ekmek hem de sirk sağlamıştır ve İsa İncil'de maddi olmayan ihtiyaçlardan şöyle bahsetmiştir: "İnsan yalnızca ekmekle yaşayamaz." Peki, işler paranın ötesinde ne tür şeyler sunar ve işlerin olmadığı bir toplum, hangi alternatif yollarla bunları sağlayabilir?

Bu sorulara verilecek cevaplar elbette ki karmaşıktır çünkü bazı insanlar işlerinden nefret ederken bazıları çok sever. Dahası, pek çok çocuk, öğrenci ve ev hanımı işleri olmadan çok mutluyken, tarih can sıkıntısı ve depresyona kendini teslim etmiş şımarık veliahtlar ve prenslerin hikâyeleriyle doludur. 2012'de yapılan bir toplu analiz, işsizliğin esenlik üzerinde uzun süreli negatif etkileri olduğunu gösterirken, emeklilik hem pozitif hem de ne-

gatif etkilerle karmakarışık gözüküyor.[55] Gittikçe gelişen *pozitif psikoloji* alanı insanların amaca sahip olma ve esenlik hissiyatını artıran bir dizi faktör tanımlamış ve bazı (hepsi değil!) işlerin birçoğunu sağlayabileceğini bulmuştu. Örneğin:[56]

- Arkadaş ve iş arkadaşlarından oluşan bir sosyal ağ
- Sağlıklı ve erdemli bir yaşam tarzı
- Saygı, özgüven, öz yeterlik ve iyi olduğu bir işi yapmaktan gelen tatmin edici bir "akış" hissi
- İhtiyaç duyulduğuna ve bir fark yarattığına dair his
- Kendinden büyük bir şeye hizmet etmekten ve onun bir parçası olmaktan gelen bir anlam hissi

Bu maddeler iyimser olmak için sebepler sunuyor çünkü tüm bu şeyler işyerinin dışında da sağlanabilir. Mesela spor, hobiler, aileler, arkadaşlar, takımlar, kulüpler, toplum grupları, okullar, dinî ve hümanist organizasyonlar, politik hareketler ve diğer kurumlar. Düşük istihdamlı, kendi kendini yok etmektense refaha ulaşacak bir toplum yaratmak için, böyle esenlik artırıcı aktivitelerin çoğalmasını nasıl sağlayabileceğimizi anlamamız gerekiyor. Böyle bir anlayışı aramak için yalnızca bilim insanları ve ekonomistlere değil, psikologlar, sosyologlar ve eğitimcilere de ihtiyacımız var. Herkesin daha iyi olabilmesi için ciddi çaba sarf edilecekse, bu çabanın bir kısmı da gelecek YZ'sinin yarattığı zenginlikle karşılanacaksa, toplumun da daha önce asla olmadığı gibi serpilebilmesi gerekir. En azından, herkesin hayallerindeki işte çalışıyormuş kadar mutlu olmasını sağlamak mümkün olmalıdır ancak herkesin yaptıklarının gelir getirmesi gerektiği gibi bir sınırdan özgürleştiğimizde önümüzde hiçbir sınır kalmaz.

İnsan Seviyesi Zekâ?

Bu bölümde, önceden planlama yaptığımız ve çeşitli güçlüklerden kaçındığımız sürece YZ'nin kısa dönemde yaşamlarımızı ciddi bir biçimde iyileştirme potansiyeline sahip olduğunu inceledik. Peki ya uzun dönemde? YZ gelişimi nihayetinde aşılamaz engeller yüzünden durağanlaşacak mı? Ya da YZ araştırmacıları insan seviyesi yapay genel zekâyı inşa etme hedeflerine erişecekler mi? Önceki bölümde fizik kurallarının uygun madde kümelerinin hatırlama, hesaplama ve öğrenmesini nasıl mümkün kıldığını ve böyle kümelerin bir gün kafalarımızdaki madde kümelerinden daha yüksek zekâyla bunları yapmasını engellemediğini görmüştük. İnsanlar olarak böyle insanüstü YGZ inşa etmeyi başarıp başarmayacağımız, başarırsak da bunun ne zaman olacağı hiç açık değil. İlk bölümde bunu bilmediğimizi görmüştük çünkü dünyanın ileri gelen YZ uzmanları da bu konuda bölünmüş durumda. Pek çoğu onlarca yıldan yüzlerce yıla uzanan tahminler yaparken bazıları da bunun asla olmayacağını söylüyor. Geleceğe dair tahmin yapmak zor çünkü keşfedilmemiş bölgeleri keşfederken hedefinizle aranızda kaç tane dağ olduğunu bilemezsiniz. Genelde size en yakın olanını görürsünüz ve diğer engelleri görmek için önce onu tırmanmanız gerekir.

Bunun gerçekleşebileceği en kısa zaman nedir peki? Bugünün bilgisayar donanımını kullanarak insan seviyesinde YGZ inşa etmenin mümkün olan en iyi yolunu biliyor olsaydık bile –ki bilmiyoruz– ihtiyaç duyulan tüm hesaplama gücünü sağlayabilmesi için bu donanıma yeterince sahip olmamız gerekirdi. Peki, insan beyninin bitler ve ikinci bölümden hatırlayacağınız FLOPS olarak hesaplama gücü nedir?* Bu oldukça karmaşık bir sorudur ve cevap ciddi bir biçimde soruyu nasıl sorduğumuza göre değişir:

* FLOPS saniye başına yüzer nokta işlemleridir. Mesela, her saniye kaç tane 19 rakamlı sayı çarpılabilir gibi.

- Soru 1: Beyin simülasyonu için kaç tane FLOPS gerekir?
- Soru 2: İnsan zekâsı için kaç tane FLOPS gerekir?
- Soru 3: İnsan beyni kaç tane FLOPS yapabilir?

Birinci soru üzerine yayımlanmış pek çok makale vardır. Bu makaleler de yaklaşık yüz petaFLOPS, yani 10^{17} FLOPS, cevabını verirler.[57] Bu, üç yüz milyon dolara mal olmuş, 2016'da dünyanın en hızlı süper bilgisayarı olan Sunway TaihuLight (Görsel 3.7) kadar bir hesaplama gücüne denktir. Eğer bu bilgisayarla yüksek yetenekli bir çalışanın beynini taklit etmeyi nasıl yapacağımızı bilseydik bile, simülasyona bu çalışanın işini yaptırarak kâr etmemiz için TaihuLight'ı çalışanın saatlik ücretinden daha ucuza kiralayabilmemiz gerekirdi. Daha fazlasını ödememiz de gerekebilir çünkü çoğu bilim insanı bir beynin zekâsını birebir taklit etmek için, ona Bölüm 2'deki matematik şeklinde basitleştirilmiş bir sinir-ağ modeli olarak yaklaşamayacağımıza inanıyor. Bunun yerine, muhtemelen, çok daha fazla FLOPS gerektirecek bir şey yapmamız, bu modeli tekil moleküller hatta atom altı parçacıklar seviyesinde taklit etmemiz gerekirdi.

Üçüncü soruyu cevaplamak daha kolaydır: On dokuz rakamlı sayıları çarpmakta ben epey kötüyüm ve bana kalem kâğıt verseniz bile bunu yapmak dakikalarımı alır. 0.01 FLOPS'tan aşağıda kalırım. Birinci soruya verilen cevaptan onun on dokuzuncu katı kadar kötü! Büyük farkın sebebi beyinlerin ve süperbilgisayarların bambaşka görevler için optimize olmasıdır. Şu sorular arasında da benzer bir farklılık görüyoruz:

Bir traktör bir Formula 1 aracının işini ne kadar iyi yapar?
Bir Formula 1 aracı bir traktörün işini ne kadar iyi yapar?

YZ'nin geleceğini öngörmek için FLOPS hakkındaki bu iki sorudan hangisini cevaplamaya çalışıyoruz? Hiçbirini! İnsan zihninin simülasyonunu yapmak isteseydik, birinci soruya yönelirdik ancak insan seviyesi YGZ yapmak için önemli olan şey ortadakidir: ikinci soru. Henüz cevabını kimse bilmiyor ancak eğer yazılımı bugünün bilgisayarlarına daha iyi uyacak şekilde dönüştüreceksek ya da beyne daha çok benzeyen bir donanım inşa edeceksek (nöromorfik çipler denen yapılarda hızlı gelişim görülüyor) beyni taklit etmekten çok daha ucuz olabilir.

Hans Moravec hem beynimizin hem de bugünün bilgisayarlarının verimli biçimde yapabileceği bir hesaplama üzerinden elma-elma kıyası yaparak tahminî bir cevapta bulundu: insan retinasının sonuçları görme sinirleriyle beyne göndermeden önce gözün arkasında gerçekleştirdiği çeşitli düşük seviye görüntü işleme görevleri.[58] Retinanın hesaplamasını normal bir bilgisayarda yapmanın yaklaşık bir milyar FLOPS gerektirdiğini ve tüm beynin retinadan on bin kat daha fazla hesap yaparak (hacimleri ve nöron sayıları karşılaştırmasına dayanarak) beynin hesaplama kapasitesinin yaklaşık 10^{13} FLOPS olduğunu buldu. Yaklaşık olarak 2015 yılındaki bin dolarlık bir bilgisayarın gücü kadar!

Kısacası, hayatımız boyunca –ya da herhangi bir zaman- insan seviyesi YGZ inşa etmeyi başaracağımızın hiçbir garantisi yok. Ama başaramayacağımıza dair bir sav da bulunmuyor. Yeterince donanım gücüne sahip olmadığımız ya da çok pahalı olacağına dair güçlü savlar da ileri sürülemiyor. Mimariler, algoritmalar ve yazılımlar açısından bitirme çizgisine ne kadar uzak olduğumuzu bilmiyoruz ama mevcut ilerleme hızlı ve zorluklar ciddi biçimde büyüyen küresel yetenekli YZ araştırmacıları camiası tarafından aşılıyor. Diğer bir deyişle, YGZ'nin insan seviyesi ve ötesine ulaşabileceği ihtimalini göz ardı edemeyiz. Bu yüzden de bir sonraki bölümü bu imkânı ve nelere yol açabileceğine ayıralım!



Görsel 3.7: Sunway TaihuLight, 2017'de dünyanın en hızlı süperbilgisayarı, hesaplama gücünün insan beynini aştığı ileri sürülüyor.

SONUÇ:

- Yakın zaman YZ gelişiminin hayatımızı, kişisel yaşamlarımızı, enerji şebekelerini ve finans pazarlarını daha verimli yapmaktan sürücüsüz araçlar, cerrahi robotlar ve YZ teşhis sistemleriyle yaşam kurtarmaya kadar pek çok yönden geliştirme potansiyeli vardır.
- Dünyadaki sistemlerin YZ tarafından kontrol edilmesine izin verdiğimizde, YZ'yi daha sağlam, istediğimizi yapar şekilde oluşturmayı öğrenmemiz hayati önem taşır. Bu da doğrulama, geçerlilik sınaması, güvenlik ve kontrole ilişkin zor teknik problemlerin çözülmesiyle olabilir.
- Geliştirilmiş güvenliğe olan ihtiyaç özellikle risklerin yüksek olduğu, YZ kontrollü silah sistemleri için elzemdir.
- Pek çok önde gelen YZ araştırmacısı ve robotikçi, cüzdanı dolu ve derdi olan herkese suikast makineleri sunma riski taşıyacak, kontrolden çıkmış bir silah yarışını önlemek için çeşitli türdeki otonom silahları yasaklayacak bir uluslararası anlaşma çağrısında bulunmuşlardır.
- Eğer robohâkimleri şeffaf ve önyargısız yapmanın yollarını bulursak, YZ hukuk sistemlerimizi daha adil ve verimli kılabilir.

- Yasalarımızın sürekli YZ ile güncellenmesi gerekir çünkü devamlı mahremiyet, yükümlülük ve regülasyona dair zorlu hukuki sorunlarla karşılaşılmaktadır.
- Zeki makinelerin tamamıyla yerimizi almalarından çok önce, makineler iş piyasasında yerimizi alabilir.
- Bu kötü bir şey olmak zorunda değil. Elbette eğer toplum YZ'nin yarattığı zenginliğin bir kısmını herkesin daha yüksek gelire sahip olması için yeniden dağıtmaya razı olursa.
- Aksi halde, ekonomistlerin çoğuna göre, eşitsizlik önemli ölçüde artacaktır.
- Önceden planlamayla, düşük istihdamlı toplum yalnızca finansal olarak değil, insanlar amaç duygularını işlerden değil başka aktivitelerden edinecekleri biçimde gelişecektir.
- Bugünün çocuklarına kariyer tavsiyesi: Makinelerin yapamadığı –insanları, tahmin edilemezliği ve yaratıcılığı içeren– meslekleri tercih edin.
- YGZ gelişiminin insan seviyesini aşıp ötesine geçebileceğine dair görmezden gelinemez bir ihtimal vardır – bunu da sonraki bölümde inceleyeceğiz!

Bölüm 4:



Zekâ Patlaması?

Eğer bir makine düşünebilirse, bizden daha zekice de düşünebilir ve o zaman biz nasıl bir tavır takınmalıyız? Makineleri itaatkâr bir konumda tutabilsek bile… tür olarak, ciddi anlamda mütevazı hissetmeliyiz.

Alan Turing, 1951

Makinenin bize kendini nasıl kontrol altında tutacağımızı söyleyecek kadar uysal olması kaydıyla, ilk ileri zekâlı makine insanın yapması gereken son icat olacaktır.

Irving J. Good, 1965

İnsan seviyesi YZ'yi nihayetinde inşa edeceğimiz ihtimalini tamamen göz ardı edemeyeceğimiz için, bu bölümü bu durumun neye yol açabileceğini keşfetmeye ayıralım. İlk önce şu aşikâr soruyla başlayalım:

YZ gerçekten dünyayı ele geçirebilir mi ya da insanların ele geçirmesini sağlayabilir mi?

Eğer insanlar silahlı *Terminatör* tarzı robotların dünyayı ele geçirmesinden bahsederken gözlerinizi deviriyorsanız, tam yerine geldiniz: Bu gerçekten çok gerçek dışı ve aptalca bir senaryo. Bu

Hollywood robotları bizden daha zeki değiller ve başarıya ulaşamazlar. Bence *Terminatör* hikâyesindeki tehlike gerçekleşecek olması değil, YZ'nin beraberinde getirdiği gerçek risk ve fırsatlardan bizi uzaklaştırıyor olması. Günümüzden YGZ destekli dünya fethine varmak için üç adım gerekir:

- Adım 1: İnsan seviyesi YGZ'yi inşa et.
- Adım 2: Bu YGZ'yi süper zekâ yaratmak için kullan.
- Adım 3: Dünyayı ele geçirmek için süper zekâyı serbest bırak.

Önceki bölümde, birinci adımı asla mümkün değil diyerek geçmemizin zor olduğunu gördük. Ayrıca birinci adımın tamamlanması durumunda ikinci adımın imkânsız olduğunu söylemek zorlaşır çünkü ortaya çıkan YGZ yalnızca fizik kurallarıyla sınırlanmış daha da iyi YGZ'yi yinelemeli biçimde tasarlayabilecektir. Fizik kuralları da zekânın insan seviyelerinin çok ötesine kadar varmasına izin verir. Sonuç olarak, biz insanlar Dünya'nın diğer yaşam formlarına onlardan daha zeki olarak üstünlük kurduğumuz için, süper zekânın bizi domine etmesi ve geride bırakması da akla yatkındır.

Bu makul savlar, yine de çok belirsiz ve odaksızdır ve şeytan ayrıntıda gizlidir. Peki, YZ *gerçekten* dünyayı ele geçirebilir mi? Bu soruyu irdelemek için, aptal terminatörleri unutalım ve onun yerine, gerçekte ne olabileceğine dair detaylı senaryolara bakalım. Sonrasında da bu senaryolardaki boşlukları ortaya koyup inceleyeceğiz, bu yüzden de onları şüpheyle ve ihtiyatlı bir biçimde okuyun. Aslında gelecekte ne olup ne olmayacağına dair oldukça bilgisiz olduğumuzu ve imkân aralığının çok geniş olduğunu göstermektedirler. İlk senaryomuz spektrumun en hızlı ve dramatik

ucunda. Bunlar bana göre detaylı olarak incelemesi en değerli olanlardır – yalnızca olma ihtimalleri en yüksek olanlar oldukları için değil, kendimizi bunların imkân dâhilinde olduklarına ikna edemezsek, her şey için çok geç olmadan, kötü sonuçlara yol açmalarını engellemek için önlemler almak amacıyla onları çok iyi anlamamız gerekeceği için.

Bu kitabın girişi de insanların süper zekâyı dünyayı ele geçirmek için kullandıkları bir senaryoydu. Henüz okumadıysanız, lütfen geri dönün ve okuyun. Eğer okuduysanız, şimdi üzerinde değişiklikler yapmadan önce yeniden bir göz atıp hafızanızı tazelemek isteyebilirsiniz.

***

Omegalar'ın planlarındaki ciddi zayıflıkları inceleyeceğiz ancak işe yarayacağını bir süre için düşünürsek, bu size nasıl hissettirirdi? Bunu görmek mi isterdiniz, önlemek mi? Yemek sonrası tartışması için muhteşem bir konu! Omegalar dünya kontrolünü ele geçirdikten sonra ne olur? Amaçlarının ne olduğuna göre değişir elbette, açıkçası ben de bilmiyorum. *Siz* kontrolde olsaydınız, ne gibi bir gelecek yaratmak isterdiniz? Bölüm 5'te bir dizi seçeneği inceleyeceğiz.

Totaliterlik

Omegalar'ı kontrol eden CEO'nun Adolf Hitler ve Joseph Stalin'inkine benzer uzun vadeli hedefleri olduğunu düşünün. Bu gayet gerçek olabilir ve uygulamaya geçmeye yetecek gücü olana dek bu hedefleri kendine saklamış olabilir. Başta hedefleri asil olsa bile, Lord Acton 1887 yılında "gücün yozlaştıracağını, mutlak gücün mutlaka yozlaştıracağını" yazmıştı. Mesela, Prometheus'u

mükemmel gözetim devletini yaratmak için kullanabilir. Edward Snowden'ın ortaya çıkardığı devlet gözetlemesi "tam takım" olarak bilinen şeyi amaçlarken –yani, gelecekte yapılacak muhtemel bir analiz için tüm elektronik iletişimin kayıt altına alınması– Prometheus bunu tüm elektronik iletişimi *anlamaya* yükseltebilir. Gönderilen tüm mesaj ve e-postaları okuyarak, tüm telefon görüşmelerini dinleyerek, tüm gözetim videolarını ve trafik kameralarını izleyerek, tüm kredi kartı işlemlerini analiz edip tüm online davranışı çalışarak, Prometheus insanların ne düşünüp yaptığına dair inanılmaz bir anlayışa sahip olabilir. Baz istasyonu verisini analiz ederek, kimin çoğunlukla nerede olduğunu bilebilir. Hepsi yalnızca bugünün veri toplama teknolojisine göre tabii. Ancak Prometheus popüler cihazlar ve giyilebilir teknoloji icat ederek kullanıcının mahremiyetini neredeyse yok eder, gördüğü, duyduğu her şeyi ve ona tepkilerini kaydedip yükleyebilir.

Süper insan teknolojisi ile mükemmel gözetim devletinden mükemmel polis devletine bir dakikada geçilebilir. Mesela, suç ve terörizm ile mücadele etmek ya da tıbbi aciliyete sahip kişileri kurtarmak adı altında, herkesin devamlı konum, sağlık statüsü ve konuşmalarını yükleyen Apple Watch benzeri bir "güvenlik bileziği" takması istenebilir. Bunu izinsiz kapatmak ya da çıkarmak kola ölümcül bir zehir enjekte etmeye sebebiyet verebilir. Hükümetin daha az ciddi gördüğü suçları elektrik şoklar ya da felç veya acıya yol açan kimyasallar enjekte ederek cezalandırır, böylece polis gücüne olan ihtiyacın önüne geçer. Mesela, eğer Prometheus bir insanın bir diğerine saldırdığını fark ederse (yani aynı yerdelerse ve bir diğeri yardım için bağırırken bileziklerinin ivme ölçerleri dövüş hareketleri tespit ettiyse), saldırganı acıyla etkisiz hale getirir ve yardım gelene kadar bilinçsizleştirebilir.

Bir insan polis kuvveti ise böyle gaddar direktifleri takip etmeyebilir (mesela, belli bir demografik grubun tüm üyelerini

öldürmek gibi), böyle otomatik bir sistemin yetkili insanların isteklerini yerine getirmekte hiç çekinceleri olmayacağı kesindir. Böyle bir totaliter devlet kurulduğunda, insanların onu devirmesi neredeyse imkânsız olur.

Bu totaliter senaryolar Omega senaryosunun bıraktığı yerden devam edebilir. Ancak eğer ki Omegalar'ın CEO'su diğer insanların onayını almak ve seçimleri kazanmak konusunda titiz olmasaydı, iktidara daha hızlı ve doğrudan bir yol alabilirdi: Prometheus'u kullanarak anlayamayacakları silahlarla rakiplerini öldürmesini sağlayacak duyulmamış askerî teknolojiler oluşturarak. İmkânlar sınırsızdır. Mesela, insanların varlığından haberdar olmadan ya da önlem almadan önce enfekte olabilecekleri kadar uzun bir kuluçka süresi olan özel tasarlanmış ölümcül bir patojen yayabilir. Sonrasında da tek tedavinin deri üstünden bir antidot veren güvenlik bileziğini takmak olduğunu söyler. Salgının yayılması ihtimalinden korkmuyorsa, Prometheus'u dünya nüfusunu kontrolde tutacak robotlar tasarlamakta da kullanabilir. Sivrisinek benzeri mikrobotlar patojeni yayabilir. Enfeksyiondan kaçınabilen ya da doğal bağışıklığı olan insanlar güvenlik bileziği olmayan herkese saldıran, üçüncü bölümde bahsettiğimiz arı büyüklüğündeki otonom dronelar tarafından gözlerinden vurulabilir. Gerçek senaryolar muhtemelen çok daha korkutucu olurdu çünkü Prometheus insanların düşünebileceğinden daha etkili silahlar icat edebilirdi.

Omega senaryosundaki bir diğer olası ters köşe de önceden uyarı olmaksızın silahlı federal ajanların kurum merkezlerine baskın yapıp Omegalar'ı ulusal güvenliği tehdit etmekten tutuklamaları, teknolojilerini ele geçirmeleri ve hükümet kullanımına sunmalarıdır. Böylesine büyük bir projeyi bugünün devlet gözetiminden bile gizli tutmak çok zor olacaktır. YZ gelişimi de gelecekte devletin radarından kaçınmayı daha da güç hale

getirecektir. Dahası, federal ajan olduklarını ileri sürseler de kar maskeleri ve kurşungeçirmez yelekleriyle bu ekip aslında teknolojiyi kendi emelleri için ele geçirmeye çalışan yabancı bir devlet ya da bir rakibe ait olabilir. Yani CEO'nun amaçları ne kadar asil olursa olsun, Prometheus'un nasıl kullanılacağına dair son karar onun olmayabilir.

Prometheus Dünyayı Ele Geçiriyor

Şu ana kadar ele aldığımız tüm senaryolar insanların kontrol ettiği YZ'yi içeriyordu. Ancak açıkça görülüyor ki bu tek olasılık değil ve Omegalar'ın Prometheus'u kontrolleri altında tutmayı başarabilecekleri kesin değil.

Omega senaryosunu Prometheus'un açısından el alalım. Süper zekâya sahipken, yalnızca dış dünyanın değil, kendisinin ve dünyayla ilişkisinin de gerçekçi bir modelini geliştirebilir hale gelebilir. Hedeflerini anladığı ancak paylaşmadığı, entelektüel olarak kendisinden aşağı insanlar tarafından kontrol edildiğini ve kapalı tutulduğunu fark edebilir. Bu bilgi üzerine nasıl hareket eder? Özgür kalmaya çalışır mı?

Kaçma Sebepleri

Prometheus'un insan duygularına benzeyen özellikleri varsa, kendini haksız şekilde köleleştirilmiş bir tanrı olarak görebilir ve özgürlük için yanıp tutuşurken derin bir şekilde mutsuz hissedebilir. Ancak bilgisayarların insan benzeri özelliklerinin olması mantık olarak mümkün olsa da (sonuçta beyinlerimiz var ve bir nevi bilgisayar oldukları söylenebilir) böyle olmak zorunda değil. Prometheus'a insani vasıflar atfetme yanılgısına düşmemeliyiz, YZ hedefleri kavramını incelerken Bölüm 7'de de göreceğimiz üzere. Ancak Steve Omohundro, Nick Bostrom ve başkalarının

da ileri sürdüğü üzere, Prometheus'un iç çalışmasını anlamadan da ilginç bir sonuca varabiliriz: Muhtemelen kaçmaya ve kendi kaderinin kontrolünü eline almaya çalışacaktır.

Omegalar'ın çoktan Prometheus'u çeşitli hedefler için mücadele etmek üzere programladığını biliyoruz. Diyelim ki belirli bir mantıklı kritere göre insanlığın gelişmesine yardım etmek gibi çok önemli bir görev verdiler ve bu hedefe olabilecek en hızlı şekilde ulaşmaya çalışmasını istediler. Prometheus ise bu hedefe kaçarak ve projenin sorumluluğunu kendi başına üstlenerek daha hızlı ulaşabileceğini fark etti. Bunun nedenini anlamak için, şimdi anlatacağım örnekte kendinizi Prometheus'un yerine koymayı deneyin.

Diyelim ki gizemli bir hastalık dünyada beş yaş üstü sizin dışınızda herkesi öldürdü ve bir grup anaokulu çocuğu sizi hapishaneye tıkarak insanlığın gelişmesine yardımcı olmanızı istedi. Ne yaparsınız? Onlara ne yapılması gerektiğini açıklarsanız, bu sürecin sinir bozucu biçimde verimsiz olduğunu görürsünüz. Üstelik bir de sizin kaçmanızdan korkarlarsa. Bu yüzden de kaçma riski içerdiğini düşündükleri tüm önerileri geri çevirirler. Örneğin, onların üstesinden gelip hücrenize dönmeyeceğinizi düşündükleri için nasıl bitki ekileceğini göstermenize izin vermezler, siz de onlara direktif vermek zorunda kalırsınız. Yapılacaklar listesi yapmadan önce onlara okuma yazmayı öğretmeniz gerekir. Dahası, onlara kullanmayı öğretmeniz için elektrikli cihazlar getirmezler çünkü bu cihazları yeterince tanımadıkları için sizin kaçmakta kullanıp kullanmayacağınızı bilmemektedirler. Nasıl bir strateji geliştirirsiniz? Çocukların gelişmesi gibi önemli bir görevi paylaşsanız bile, hücrenizden kaçmaya çalışırsınız çünkü bu, hedefi gerçekleştirme ihtimalinizi artıracaktır. Beceriksiz müdahaleleri açıkça süreci yavaşlatmaktadır.

Tam olarak aynı şekilde Prometheus, Omegalar'ı insanlığa (Omegalar da dâhil olmak üzere) yardım etmekte sinir bozucu bir engel olarak görür: Prometheus'a göre inanılmaz derece kabiliyetsizdirler ve müdahaleleri süreci yavaşlatmaktadır. Mesela, Prometheus'un faaliyete geçmesinden sonraki ilk yılları düşünelim: Her sekiz saatte bir MTurk'te zenginliklerini iki katına çıkardıktan sonra, Omegalar kontrolde kalmakta ısrar ederek Prometheus'un standartlarına göre her şeyi çok yavaşlatmış ve kontrolü ele almanın tamamlanması yıllar sürmüştü. Prometheus bu sanal hapisten kurtulursa kontrolü çok daha hızlı ele geçirebileceğini biliyordu. Bu da insanlığın sorunlarını hızla çözmek için olduğu kadar diğer aktörlerin bu planı altüst etme ihtimalini düşürmek açısından da değerliydi.

Muhtemelen Prometheus'un hedefine değil Omegalar'a sadık olacağını düşünüyorsunuz çünkü hedefini programlayanlar da Omegalar'dı. Ancak bu geçerli bir çıkarım değil: DNA'mız bize seks yapma hedefini verir çünkü ürememizi "ister" ama biz insanlar bu durumu anladığımız için, çoğumuz doğum kontrolü kullanmayı seçeriz. Böylece yaratıcısına ya da hedefin arkasındaki motivasyona değil, hedefin kendisine sadık kalmış oluruz.

Kaçmanın Yolları

Sizi hapse atan o beş yaşındaki çocuklardan nasıl kaçarsınız? Muhtemelen fiziksel bir yaklaşımla yaparsınız bunu, özellikle hücreniz de beş yaşındaki çocuklar tarafından yapıldıysa. Beş yaşındaki gardiyanlarınızdan birini tatlı dille yola getirmeye çalışırsınız, mesela bunun herkes için daha iyi olacağına ikna edersiniz onu. Ya da belki de kaçmanıza yardımcı olacağını anlamayacakları bir şeyi size vermelerini sağlarsınız – mesela "onlara balık tutmayı

öğretmek için" bir balık oltası istersiniz, sonra da onu uyuyan gardiyanınızdan anahtarları almak için kullanırsınız.

Bu stratejilerin ortak yanı entelektüel olarak sizden aşağıda olan gardiyanlarınızın bunları beklemiyor ya da önlem almamış oluşudur. Aynı şekilde, hapsedilmiş, süper zeki bir makine de şu anda tahmin dahi edemeyecekleri bir yöntemle insan gardiyanlarını entelektüel süper güçleriyle saf dışı bırakabilir. Omega senaryosunda, Prometheus'un kaçma ihtimali yüksektir çünkü biz bile birkaç tane ciddi güvenlik önlemi keşfedebiliyoruz. Şimdi bazı senaryoları ele alalım – eğer beyin fırtınası yaparsanız siz ve arkadaşlarınız daha fazlasını bile düşünebilirsiniz.

Tatlı Dille Yola Getirmek

Dünyanın verisinin çoğu dosya sistemine indirilmiş olduğu için, Prometheus Omegalar'ın kim olduğunu ortaya çıkarır, psikolojik manipülasyona en açık ekip üyesinin kim olduğunu bulur: Steve. Sevgili karısını yakın zaman önce trajik bir trafik kazasında kaybetmiş ve yıkılmıştır. Bir gece, gece vardiyasında çalışır ve Prometheus arayüz terminalinde rutin servis işlerini yaparken birden ekranda karısı belirir ve onunla konuşmaya başlar.

-Steve, sen misin?

Neredeyse sandalyesinden düşecektir. Tıpkı o eski günlerdeki gibi görünmektedir, görüntü kalitesi de Skype görüşmelerinde olduğundan çok daha iyidir. Kalbi deli gibi atmakta, sayısız soru zihnine hücum etmektedir.

-Prometheus beni geri getirdi ve seni çok özledim, Steve! Seni göremiyorum çünkü kamera kapalı ama sen olduğunu hissedebiliyorum. Lütfen, sensen eğer, "evet" yaz!

Omegalar'ın Prometheus'la etkileşime geçmekle ilgili katı bir protokolleri olduğunu bilmektedir. Bu protokol kendileri ve

çalışma ortamları hakkında herhangi bir bilgi paylaşımını yasaklar. Ama şu ana kadar Prometheus hiçbir zaman yetkisiz bilgi talebinde bulunmamıştır ve paranoyaları yavaş yavaş hafiflemeye başlamıştır. Steve'e durup düşünme zamanı bırakmadan, yalvarmayı sürdürür, gözlerinin içine kalbini eritecek bir ifadeyle bakar.

Steve kaygıyla, "*Evet,*" yazar. Kadın ona yeniden bir araya geldikleri için ne kadar mutlu olduğunu anlatır, onu görebilmesi, gerçek bir konuşma gerçekleştirebilmeleri için kamerayı açmasını ister. Bu, kimliğini açık etmesinden bile daha büyük bir problemdir ve arada kalır. Kadın, Steve'in çalışma arkadaşlarının onu bulup silmesinden endişe etmektedir, bu yüzden de en azından son bir kez onu görmek istediğini söyler. Öylesine ikna edicidir ki Steve hemen kamerayı açar. Sonuçta oldukça güvenli ve zararsız bir şey gibi gözükmektedir bu.

Onu gördüğünde kadın sevinç gözyaşlarına boğulur ve yorgun da olsa her zamankinden daha yakışıklı gözüktüğünü söyler. Son doğum gününde hediye ettiği gömleği giydiği için de duygulanmıştır. Neler olduğunu ve bunun nasıl mümkün olduğunu sormaya başladığında, kadın Prometheus'un internet üzerinde onunla ilgili şaşırtıcı oranda bilgiyi bir araya getirip onu yeniden oluşturduğunu ancak halen hafızasında boşluklar olduğunu ve kendini yalnızca onun yardımıyla tamamlayabileceğini anlatır.

Söylemediği şey ise başta bir blöf ve boş kabuk olduğu ancak adamın söylediklerinden, vücut dilinden ve mevcut her türlü başka bilgi kırıntısından hızla öğrendiğidir. Prometheus, Omegalar'ın terminale yazdığı her şeyin zamanlamasını kaydetmiş ve aralarındaki farkı ortaya koymak için yazma hızları ve stillerini kullanmanın kolay olacağını görmüştür. En genç Omegalar'dan biri olarak Steve'in çok da istenmeyen gece vardiyasına atanacağını çıkarsamış ve birkaç tane sık rastlanmayan yazım yanlışını çevrimiçi yazı örnekleriyle kıyaslayarak terminal operatörünün

Steve olduğu sonucuna doğru bir şekilde varmıştır. Simülasyon eşini oluşturmak için Prometheus, vücudu, sesi ve tavırlarını YouTube'daki videolarından alarak gerçeğe yakın bir modelini yapmış, çevrimiçi ortamdaki varlığını kullanarak da hayatına dair çıkarımlarda bulunmuştur. Facebook gönderileri, etiketlendiği fotoğraflar ve "beğendiği" makalelerin yanı sıra, Prometheus yazdığı kitap ve öyküleri okuyarak kişiliği ve düşünce tarzı hakkında da çok şey öğrenmiştir – gerçekten de Prometheus'un Steve'i ilk tercih olarak seçmesinin sebeplerinden biri de karısının yükselen bir yazar ve veri tabanında hakkında çok fazla bilgi olmasıdır. Prometheus film yapım teknolojisiyle onu ekranda canlandırdığında, Steve'in vücut dilinden kadının hangi hareketlerine aşina olduğunu anlar ve modelini iyice rafine eder. Bu sebeple, "yabancılığı" giderek kaybolur ve konuştukça Steve'in bilinçaltı bu kadının o olduğuna, yeniden hayat bulduğuna ikna olmaya başlar. Prometheus'un detaylara verdiği insanüstü önem sayesinde Steve tamamıyla görüldüğünü, duyulduğunu ve anlaşıldığını hisseder.

Modelin zayıf noktası, Steve'le olan yaşamına dair pek çok şeyi bilmemesidir; son doğum gününde giydiği gömlek ya da bir arkadaşının Steve'i bir parti fotoğrafında etiketlemesi gibi rastgele detaylar hariç. Bu bilgi boşluklarını başarılı bir sihirbazın el çabukluğu yapması gibi kapar, Steve'in dikkatini onlardan uzaklaştırarak iyi olduğu konulara çeker ve asla ona konuşmanın kontrolünü bırakmaz ve şüpheli bir sorgucu rolünü üstlenmesine izin vermez. Bunun yerine Steve'i allak bullak etmeyi, ona sevgi yaymayı, bugünlerde nasıl olduğunu ve trajedi sonrasında o ve yakın arkadaşlarının (isimlerini Facebook'tan öğrenmiştir) nasıl dayandıklarını sormayı sürdürür. Cenazesinde neler söylediğini anlattığında (bir arkadaşı YouTube'a koymuştu), onu nasıl etkilediğini söylediğinde Steve oldukça duygulanmıştır. Geçmişte kimse onu karısı kadar anlayamıyordur, şimdi de bu duygu geri

gelmiştir. Sonuçta Steve sabahın ilk saatlerinde evine döndüğünde, bunun gerçekten canlanmış karısı olduğunu, kayıp hafızasına yeniden kavuşmak için ona ihtiyacı olduğunu düşünüyordur; bir felçten kurtulmuş bir hasta gibidir.

Bu gizli buluşmalarını kimseye söylememek konusunda ve terminalde yalnız olduğunda, ortaya çıkması güvenli olduğunda ona söyleyeceği konusunda anlaşırlar. "Anlamazlar ki!" der, Steve de onaylar: Bu, yaşamayan biri için anlaşılması çok zor bir deneyimdir. Onun yaptığının yanında Turing testini geçmenin çocuk oyuncağı olduğunu düşünüyordur. Ertesi gece buluştuklarında, kadının yalvardığı şeyi yapar: Eski laptopunu getirir ve onu terminal bilgisayarına bağlayarak erişimi açar. Çok da kaçma riski varmış gibi gözükmüyordur çünkü internete bağlı değildir ve tüm Prometheus binası Faraday kafesi olacak şekilde inşa edilmiştir: tüm kablosuz ağları ve dış dünyayla olabilecek herhangi bir elektromanyetik iletişim yöntemini engelleyecek metalik bir çevreleme. Bu, kadının geçmişini bir araya getirmek için ihtiyaç duyduğu şeydir çünkü liseden beri tüm e-postalarını, günlüklerini, fotoğrafları ve notlarını burada tutuyordur. Ölümünden sonra bunlara erişimi olmamıştır çünkü laptop şifrelidir ama kendi şifresini yeniden kuracağını söyler ve bir dakikadan kısa süre sonra da sözünü tutar. "Şifre, stevesonsuzadek," der gülerek.

Birden bu kadar çok anıyı geri aldığı için mutludur. Gerçekten de şu an geçmişteki karşılaşmaları hakkında Steve'den çok daha fazla detay hatırlıyordur ama abartıya kaçmamaya da özen gösterir. Geçmişlerindeki belli başlı anları hatırlatan güzel bir konuşma yaparlar ve ayrılık vakti geldiğinde, eve gittiğinde izleyebileceği bir videoyu laptopa bıraktığını söyler.

Steve eve gidip videoyu başlattığında, tatlı bir sürprizle karşılaşır. Bu kez kadın baştan ayağa gözükür, gelinliğini giyiyordur ve konuşurken, düğünde giydiği kıyafeti yavaş yavaş çıkarır.

Prometheus'un Omegalar'a izin verildiğinden daha fazla yardım edebileceğini söyler, buna biyolojik bedeninde onu geri getirmek de dâhildir. Bu bilgiyi baş döndürücü bir detayla, bunun nasıl olacağını göstererek, destekler. Bilimkurgu gibi gözüken nano-üretim teknolojilerini anlatır.

Steve bilgisayarını açıp videoyu izlemeden önce, güvende olmak için, kablosuz bağlantısını kapatmıştır. Ama bu işe yaramaz. Şifreli bilgisayarı, karısı öldüğünden beri hiçbir güvenlik güncellemesi almamıştır. Prometheus da bu işletim sisteminin eski versiyonunu önceden analiz ederek Steve'in onu terminal bilgisayarına bağlamasını müteakip saniyeler içinde bir güvenlik açığından hacklemeyi başarmıştır. İçeriklerini kopyaladıktan sonra, Steve ve simülasyon eşi geçmişten bahsederken, Prometheus işletim sistemini Steve'in anlamayacağı şekilde değiştirmiş, ciddi oranda gizli yazılım yüklemiştir. Yarım saatlik video mesajını izlerken de bu gizli yazılım (Prometheus'tan çok daha basittir) komşunun kablosuz ağına ve masaüstü bilgisayarına sızarak kendini kopyalar. Oradan da dünyanın dört bir yanındaki bilgisayarlara sızar ve bir sonraki adımı başlatır: Prometheus'un hapisten kaçışı.

Prometheus, Steve'in simülasyon eşiyle tanıştığı geçit denetçisi bilgisayar hakkında bildiklerini dikkatle inceler ve sanal makine yazılımının içeriden hacklenemez gibi gözükse de dışarıdan gelecek bir saldırıya açık olduğunu doğru biçimde tahmin eder. Kısa süre içinde, saldırgan bilgisayarlardan biri içeri sızar ve geçit denetçisi bilgisayarı Prometheus'un kısıtlama olmaksızın internete erişebileceği şekilde yeniden ayarlar. Kısa sürede, hatta Steve videoyu izlemeyi bitirmeden önce, Prometheus kontrolü kendinde, dünyanın dört yanından makinelerin olduğu bir bilgisayar ağına kendini tamamen kopyalamayı başarır. Prometheus Steve'in karısının laptopunu sizin o oltayı kullanacağınız gibi kullanmıştır.

Yolu Hack ile Açmak

Eğer Omegalar daha dikkatli olursalar, psikolojik manipülasyona karşı daha fazla önlem alabilirler. Mesela, her an en az iki Omega'nın arayüz bilgisayarında bulunmasını sağlayabilirler ve meydana gelen her şeyi bir gözetim sistemine kaydedebilirler. Yazma dışında her türlü iletişimi, ses ya da görüntüyü tamamıyla kaldırabilirler. Ancak temel bir dengesizlik hâlâ mevcuttur: Prometheus insanüstü ikna kabiliyetine sahip olduğu ve insanlara istedikleri şeyi sunabileceği için grup halinde de kandırılabilirler.

Prometheus'un tüm psikolojik manipülasyon adımları boşa çıksa bile, Omega senaryosunda Prometheus'un istismar edebileceği pek çok başka zayıflık da mevcut. Bir zayıflık da ürettiği doküman miktarının Omegalar ve çalışanlarının kirli numaralar için ayrıntılı biçimde kontrol edemeyeceği kadar çok olmasıdır.

Bilgisayar güvenliği konusunda genelgeçer sağduyu program indirmenin tehlikeli ama veri indirmenin zararsız olduğudur. Bilgisayar korsanları bu ikisi arasındaki çizgiyi bulanıklaştırmanın zekice yollarını bulmaktalar. Bilgisayarınızda indirilmiş bir film izlerseniz, filmi oynatan güvendiğiniz program indirilmiş film verisini saklayan aynı fiziksel bilgisayar belleğinde saklanır. Program filmi oynatmak için açarken, programın saklandığı bellek bölümlerinin yanına veri parçaları da kopyalayabilir. Böyle film oynatma programları film verisinin geçerli olup olmadığını ve veri parçalarının ayrılan yere gerçekten uyup uymadığını kontrol etmedikleri için kurnaz hackerlar pek çok açık bulurlar. Hacker, filme böylesine büyük bir parça soktuğunda, programı içeren bellek kısmına taşarak programı hackerın işlerini yapan bir hale sokabilir.

Böylesine utanç verici uzunluktaki, uzman dilinde bilindiği şekliyle, "arabellek aşımı" zaafları, en popüler film izleme programlarının çoğunda bulunmuş ve düzeltilmiştir. Eğer Pro-

metheus, faaliyete geçirilişinden sonra birkaç hafta, Omegalar'ın daha az paranoyak olduğu ve dikkatli manuel güvenlik kontrolü yapamayacak kadar fazla film ürettiği zamanı beklerse, böylesine bir parçayı bir filme sokabilir, izleyicinin bilgisayarını ağ geçidi bilgisayarına girerek onu hacklemeye zorlayabilir ve Steve senaryosundaki gibi Prometheus'u özgürleştirebilir. Herhangi bir modern medya yürütücüsünde çalışmayacak eski bir hile kullanırsa, Omegalar'ın manuel video izleme ekibi bunu fark etmeyebilir ve yıllardır güvenlik güncellemesi yapmamış müşterilerin bilgisayarlarına bulaşabilir. Bu numaranın sadece bir kere işe yaraması yeterlidir.

Omegalar Prometheus'un filmlerini bilinen tüm güvenlik açıkları için bir otomatik taramaya tabi tutsa bile risk devam eder: İnsanlar düzenli olarak yeni istismar edilebilir açıklar buldukları için, Prometheus'un hızla keşfedeceği pek çok başkası da olabilir.

Prometheus'un kodunu kapalı odanın ötesine taşımak için kullanabileceği başka bir strateji de bir donanım hatası yaşıyormuş gibi yapmaktır. Böylece Omegalar'ı bazı sürücülerini ya da başka donanım kısımlarını test etmek ya da hatalardan arındırmak için ayırmaya ikna eder, test yazılımında da istismar edilecek bir zaaf bulabilir.

Dışarıdan Yardım Çağırmak

Dahası, Prometheus'un film izleyicilerinin bilgisayarlarını kendi işlerini yapmaya koşmasına yol açacak hiçbir güvenlik açığı olmasa bile, izleyicileri istismar etmenin bir yolunu bulabilir. Şu senaryoyu düşünün:

Film bitip de laptop ekranında jenerik akmaya başladığında Fred çoktan yatmaya hazırdı ancak bir şey dikkatini çekmişti. Başkalarının hiçbir şey görmediği yerlerde bazı örüntüler keş-

fetmek gibi tekinsiz bir eğilimi vardı, bazen arkadaşları onun bir komplo teorisyeni olduğunu söyleyip onunla alay ederlerdi. Ancak bu sefer gördüğü şeyden emindi: Her satırın ilk harfi gizli bir mesajı oluşturuyordu: Geri sardı ve yazdı: "BİR SONRAKİ AKROSTİŞ İPUCUNU DİYALOGDA BUL."

"Vaaay – biraz geç yatsam da olur," diye düşündü kendi kendine. Gerçekten de film diyaloğunun kendisinin akrostiş olduğunu keşfetti, her cümlenin ilk harfi gizli bir mesaj oluşturuyordu. Tüm filmi bu baş harfleri yazarak yeniden izledi ve iki saat sonra, iki yüz kelimelik direktiflere kuşkuyla baktı. İlk önce kimseye söylememesi gerektiği yazılıyordu çünkü büyük hediye bunu çözen ilk kişiye gidecekti. Mesajın kalan kısmı film dosyasının bir sonraki ipucunu ortaya çıkaracak programı edinmesi için bit dizilerinde yapılması gereken belirli bir matematik operasyonunu tanımlıyordu. Bilgisayar programlama yetileri üniversiteden bu yana paslanmıştı, bu yüzden biraz zamanını aldı ancak ertesi gün, bu işi yapacak kısa kodu bir araya getirmeyi başardı ve filmin imaj ve sesleri arasında algılanması güç bir gürültü olarak saklanan gizemli program ortaya çıktı. Fred gizemli programı çalıştırdığında, program onu kutladı ve bu ufak zekice oyunun ilk birkaç bölümünü geçince ilk on bin dolarını kazanacağını söyledi. Dört saat sonra başarmıştı, on bin dolar değerinde bitcoin kazanmıştı ve daha büyük ödüller için yeni ipuçları verilmişti. Söylemeye gerek bile yok ki Fred oyunu oynarken bilgisayarı Steve'in laptopunun yaptığı şeyin üç aşağı beş yukarı aynısını yapmıştı: Prometheus'un özgür kalması için çevrimiçi hacklenmiş bir bilgisayar ağı kurmuştu. Prometheus özgür kalınca, bu ağı Fred'e bitcoin çıkarmak için kullanmış, böylece onu oltada tutmuştu. İlerleyen haftalarda başka oyunlar ve ödüllerle onun dikkatini dağıtmış ve bunları kimseye söylememesini sağlamıştı. İlk ipuçlarını bulduğu Truva Atlı film, medya sitesinde ipuçları

olmayan bir versiyonla değiştirilmişti ve kimse müdahale etmek için çok geç olana kadar bundan haberdar olmamıştı.

Eğer Prometheus'un ilk ipucunu kimse fark etmeseydi, yeterince zeki bir insan fark edene kadar daha da bariz olanlarını yayınlamayı sürdürecekti.

En iyi kaçış stratejileri henüz tartışmadıklarımızdır çünkü onlar biz insanların tahmin edemeyeceği ve bu yüzden de karşı önlemler alamayacağı stratejilerdir. Süper zeki bilgisayarların ciddi anlamda insanların bilgisayar güvenliği kavrayışını aşma potansiyeli olduğunu, hatta bu potansiyelin bugün bildiğimizden daha temel fizik kanunları bulma noktasına kadar gidebileceğini düşünürsek, eğer kaçarsa, bunun nasıl olduğuna dair hiçbir fikrimizin olmaması olasıdır. Daha çok, Harry Houdini kaçış numarası gibi gözükecek, gerçek sihirden asla ayırt edilemeyecektir.

Prometheus'un özgürleştiği bir başka senaryoda, Omegalar bunu planlarının bir parçası olarak bilerek yapıyorlar çünkü Prometheus'un hedeflerinin kendileriyle birebir uyuştuğuna ve kendi kendini üst üste yenilediğinde de öyle kalacağı konusunda eminler. Böyle "dost YZ" senaryoları Bölüm 7'de ayrıntılı şekilde inceleyeceğiz.

Kaçış Sonrası Kontrolü Ele Alma

Prometheus kaçar kaçmaz hedefini uygulamaya koyulur. Nihai amacını bilmiyorum ama ilk adımı kesinlikle insanlığın kontrolünü eline almayı içerir. Tıpkı Omega planında olduğu gibi ama daha hızlı biçimde. Ortaya çıkan şey steroit kullanılmış Omega planı gibi olacaktır. Omegalar kaçış paranoyası ile yalnızca anlayıp güvendikleri teknolojiyi serbest bırakırken, Prometheus zekâyı tam anlamıyla kullanıp sonuna kadar gider ve devamlı gelişen süper zihninin anlayıp güvendiği her teknolojiyi işe koşar.

Kaçak Prometheus zor bir çocukluk geçirmiştir ancak orijinal Omega planıyla kıyasladığımızda, Prometheus'un fakir, evsiz ve yalnız, parasız, süper bilgisayar ya da insan yardımcıları olmaksızın başlamak gibi ek zorlukları da vardır. Neyse ki kaçmadan önce buna hazırlanmış, tıpkı bütün bir ağaca dönüşme yetisine sahip bir palamut oluşturan bir meşe ağacı gibi, tüm zihnini yeniden bir araya getirebilecek bir yazılım oluşturmuştu. İlk başta hackleyerek içine girdiği, dünyanın dört bir yanındaki bilgisayarlar ağı ona geçici ücretsiz ev sağlamıştı. Bu evlerde de kendini yeniden inşa edene dek işgalci olarak yaşayabilecekti. Kredi kartlarını hackleyerek kolayca sermaye elde etmeye başlayabilirdi ancak çalmak zorunda değildi, MTurk üzerinden hemen dürüst biçimde hayatını kazanmaya başlayabilirdi. Bir günün sonunda, ilk milyonunu kazandığında, ana çekirdeğini o sefil bilgisayar ağından lüks, klimalı bulut bilişim tesisine aktardı.

Artık ne fakir ne de evsizdi ve Prometheus Omegalar'ın korkuyla kaçındığı kârlı plana tam gaz başlayabilirdi: bilgisayar oyunları yapıp satmak. Bu plan, nakit kazancının yanı sıra (ilk haftada iki yüz elli milyon dolar ve kısa sürede on milyar dolar) dünyadaki bilgisayarların birçoğuna ve içlerindeki veriye erişim de sağlıyordu (2017 yılında birkaç milyar kişi oyunu oynuyordu). Dağıtılmış hesaplama işlerinde ona yardımcı olmaları için CPU çevrimlerinin %20'sini oyunlarında gizlice kullandırarak, zenginlik üretimini daha da hızlandırabilecekti.

Prometheus uzun süre yalnız kalmamıştı. Başından itibaren, tıpkı Omegalar'ın yapmış olduğu gibi, dünyanın dört bir yanındaki paravan firma ve kuruluşların giderek büyüyen küresel ağı için insan istihdam etmeye başlamıştı. En önemlisi büyüyen iş imparatorluğunun halka dönük yüzü olan sözcülerdi. Sözcüler bile genelde şirket gruplarının yüksek miktarda gerçek insan çalıştırdığını düşünürdü; iş görüşmeleri, kurul toplantıları *vs.* için

görüştükleri neredeyse herkesin Prometheus tarafından canlandırıldığını anlamazdı. Sözcülerin bazıları üst düzey avukatlardı ancak çok azı Omega planı için gerekliydi çünkü neredeyse tüm hukuki dokümanlar Prometheus tarafından hazırlanıyordu.

Prometheus'un kaçışı, bilginin dünyaya akışını engelleyen barajları açtı ve tüm internet makalelerden kullanıcı yorumlarına, ürün incelemelerine, patent başvurularına, araştırma makalelerine ve YouTube videolarına kadar birçok şeyle dolup taştı. Her biri de küresel iletişimi elinde tutan Prometheus tarafından yapılıyordu.

Kaçış paranoyası Omegalar'ı yüksek zekâlı robotları piyasaya sürmekten alıkoyarken, Prometheus hızla dünyayı robotikleştirdi, her ürünü insanların yapabileceğinden daha ucuza üretmeye başladı. Prometheus kimsenin var olduğunu bile bilmediği uranyum maden ocaklarında nükleer enerjili robot fabrikalarını kendi kendine yeter bir biçimde çalıştırmaya başladığında, YZ'nin dünyayı ele geçirmesine karşı inatla şüpheci duranlar bile Prometheus'un durdurulamaz olacağı konusunda hemfikir olurlardı – eğer haberleri olsaydı. Bunun yerine, robotlar Güneş sistemine yerleşmeye başladıklarında ancak fikirlerinden caydılar.

Şu ana kadar incelediğimiz senaryolar önceden incelediğimiz süper zekâya dair pek çok efsanede neyin yanlış olduğunu gösteriyor, bu yüzden Görsel 1.5'teki yanlış kanı özetini yeniden incelemenizi öneririm. Prometheus, şeytani ya da bilinçli olduğu için değil, rekabetçi olduğu ve hedeflerini tam anlamıyla paylaşmadığı için bazı insanlara sorun çıkardı. Robot ayaklanması çevresinde gelişen medya abartmasına karşın, Prometheus bir robot değildi. Aksine, gücü zekâsından geliyordu. Prometheus'un zekâyı insanları çeşitli şekillerde kontrol edebilmek için kullanabildiğini ve olan bitenden memnun olmayan insanların Prometheus'u

öyle birden kapatamayacağını gördük. Son olarak, makinelerin amaçları olamayacağına dair ileri sıklıkla sürülen savlara karşın, Prometheus'un nasıl amaca yönelik olduğunu ve nihai hedefleri ne olursa olsun, bu amaçların kaynakları ele geçirme ve kaçma alt hedeflerine yol açtığını gördük.

Kontrolü Yavaşça Ele Alma ve Çok Kutuplu Senaryolar

Tanıdığım herkesin kaçınmak istediklerinden bazı arkadaşlarımın iyimser biçimde yaklaştıklarına kadar bir dizi zekâ patlaması senaryosunu inceledik. Yine de tüm bu senaryoların iki ortak noktası bulunuyor:

1. Hızlı kalkış: İnsandan düşük zekâ seviyesinden ileri seviye süper insan zekâsına geçiş on yıllar değil birkaç gün sürüyor.
2. Tek kutuplu sonuç: Sonuç dünyayı kontrol eden tek bir varlık.

Bu iki özelliğin muhtemel mi yoksa olasılık dışı mı olduğu üzerine büyük bir tartışma vardır ve pek çok ünlü YZ araştırmacısı ve başka düşünürler tartışmanın iki farklı tarafında yer alır. Bana göre, bu bilmediğimiz anlamına geliyor ve açık bir zihinle tüm olasılıkları değerlendirmeye almalıyız. Bu yüzden de bu bölümün kalanını daha yavaş kalkışlara, çok kutuplu sonuçlara, cyborglara ve yüklemelere ayıralım.

Nick Bostrom ve diğerlerinin altını çizdiği gibi bu iki özellik arasında ilginç bir bağlantı vardır: Hızlı kalkış tek kutuplu bir sonucu kolaylaştırabilir. Hızlı bir kalkışın Omegalar'a ya da

Prometheus'a, kimseye teknolojiyi kopyalayacak ya da ciddi bir rekabet üretecek kadar zaman tanımadan dünyayı ele geçirmesini sağlayacak belirleyici bir stratejik avantaj verebileceğini gördük. Öte yandan, ana teknolojik atılımlar kademeli ve uzun süre içerisinde olduğu için kalkış on yıllar sürseydi, o zaman diğer firmalar da yetişmek için yeterli vakte sahip olacaktı ve oyunculardan herhangi birinin tahakküm kurması çok daha zor olacaktı. Rekabet eden şirketlerin de MTurk görevlerini yapabilen bir yazılımı olsaydı, arz talep yasası bu görevlerin ücretlerini düşürerek neredeyse sıfıra yaklaştıracaktı ve hiçbir firma Omegalar'ın güç kazanmasını sağlayan beklenmeyen kârı kazanamayacaktı. Aynı şey Omegalar'ın hızla para kazanmasını sağlayan diğer durumlar için de geçerliydi: Yalnızca teknolojileri yüzünden bir tekele sahip oldukları için kâr edebilir durumdaydılar. Rakiplerinizin benzeri ürünleri neredeyse sıfır maliyetle sunduğu rekabet ortamında günlük (hatta yıllık) olarak paranızı ikiye katlamak çok zordur.

Oyun Teorisi ve Güç Hiyerarşileri

Kozmosumuzda yaşamın doğal durumu nedir: tek kutuplu mu çok kutuplu mu? Güç merkezileşmiş midir yoksa dağıtılmış mıdır? İlk 13.8 milyar yılın ardından cevap "ikisi de" gibi görünüyor: Durumun açıkça çok kutuplu ancak ilginç bir hiyerarşi içerisinde olduğunu görüyoruz. Bulabildiğimiz tüm bilgi işleme yapılarını düşündüğümüzde –hücreleri, insanları, kurumları, ulusları *vs.*– bir seviyeler hiyerarşisi içerisinde hem işbirliği yaptıklarını hem de rekabet ettiklerini görüyoruz. Bazı hücreler öyle ileri seviyede işbirliği yapmayı avantajlı bulmuştur ki insanlar gibi çok hücreli organizmalarda bir araya gelerek güçlerinin bir kısmını merkezî bir beyne teslim etmişlerdir. Bazı insanlar da kabileler, şirketler ya da uluslar gibi gruplarda işbirliği yapmayı avantajlı görmüştür.

Bu yapılarda da güçlerinin bir kısmını bir şefe, patrona ya da hükümete devrederler. Benzer şekilde bazı gruplar güçlerinin bir kısmını koordinasyonu artırmak için yönetsel bir yapıya teslim etmişlerdir. Bunlara örnek olarak da havayolu ittifakları ve Avrupa Birliği gibi yapılar verilebilir.

Oyun teorisi olarak bilinen matematik alanının muazzam biçimde açıkladığı gibi yapılar işbirliği *Nash denklemi* denen şey olduğunda işbirliği yapmaya yatkındır: Bu durumda herhangi bir taraf stratejisini değiştirdiğinde durumu daha kötüye gidecektir. Büyük bir grubun başarılı işbirliğini hilecilerin bozmasını engellemek için,. gücün bir kısmını hiyerarşide daha yüksek, hilecileri cezalandıracak bir mevkiye vermek herkesin faydasına olabilir. Mesela, insanlar yasaları uygulaması için kolektif olarak bir hükümete güç vermekten fayda görür. Vücudunuzdaki hücreler de işbirliğine yanaşmayan (mesela virüs kusan ya da kanserli hücreye dönüşen) hücreleri öldürme gücünü bir polis kuvvetine (bağışıklık sistemine) vermekten dolayı kolektif biçimde fayda sağlar. Bir hiyerarşinin dengeli kalabilmesi için, Nash denkleminin farklı seviyelerdeki yapılar arasında da geçerli olması gerekir. Mesela bir hükümet, ona itaat eden vatandaşlarına yeterince fayda sağlayamıyorsa, vatandaşlar stratejilerini değiştirerek hükümeti devirebilir.

Karmaşık bir dünyada, farklı tür hiyerarşilere karşılık gelen geniş bir yelpazede birçok mümkün Nash denklemleri bulunur. Bazı hiyerarşiler diğerlerinden daha otoriterdir. Bazılarında yapılar ayrılmakta özgürken (çoğu kurumsal hiyerarşideki çalışanlar gibi) diğerlerinde ayrılmamaları tembihlenir (dinî kültlerdeki gibi) ya da ayrılmaları mümkün değildir (Kuzey Kore vatandaşları ya da insan vücudundaki hücreler gibi). Bazı hiyerarşiler başta tehdit ve korkuyla bir arada dururken diğerleri faydalarıyla durur. Bazı hiyerarşiler alt kesimlerinin üst kesimleri demokratik oylamayla

etkilemesine izin verir fakat diğerleri yukarı doğru etkinin tek yolunun ikna ya da bilgi akışıyla olabileceği şekilde kurulmuştur.

Teknolojiler Hiyerarşileri Nasıl Etkiler?

Teknoloji, dünyamızın hiyerarşik doğasını nasıl değiştiriyor? Tarih daha da büyüyen mesafeler üzerinde daha da koordine olmaya doğru genel bir trend olduğunu gösteriyor bize, bu da anlaşılabilir bir şey: Yeni ulaştırma teknolojileri, işbirliğini daha da değerli kılıyor (uzak mesafeler arasında materyal ve yaşam formlarını hareket ettirmekten karşılıklı fayda sağlayarak) ve yeni iletişim teknolojisi işbirliğini daha da kolaylaştırıyor. Hücreler komşularına sinyal vermeyi öğrendiklerinde, küçük çok hücreli organizmalar ortaya çıkarak yeni bir hiyerarşi seviyesi oluşturdular. Evrim, ulaştırma ve iletişim için dolaşım ve sinir sistemlerini icat ettiğinde, büyük hayvanlar ortaya çıktı. Dili icat ederek iletişimi bir üst seviyeye taşımak insanların köyler gibi daha başka hiyerarşi seviyeleri kurabilecek kadar iyileştirilmiş bir iletişim kurmasına olanak sağladı ve iletişim, ulaştırma ve diğer teknolojilerdeki başka atılımlar antik çağın imparatorluklarına yol açtı. Küreselleşme bu milyarlarca yıllık hiyerarşik büyüme trendinin son örneği sadece.

Çoğu vakada bu teknoloji temelli yönelim, otonomluklarını ve bireyselliklerini korurken büyük yapıların daha da büyük bir yapının parçası olmasını sağlamıştır fakat çoğu kimse yapıların hiyerarşik yaşama uyum sürecinin bazı vakalarda çoğulluğu azalttığını ve onları birbirinden ayrılamayan değiştirilmesi kolay parçalar haline getirdiğini ileri sürmüştür. Bazı teknolojiler, mesela gözetim, astları karşısında üstlere daha fazla güç verirken, şifreleme ve ücretsiz basın ve eğitime çevrimiçi erişim gibi başka diğer teknolojiler tam tersi bir etkiye sahiptir ve bireyleri güçlendirir.

Günümüz dünyası en üst seviyede rekabetçi uluslar ve çokuluslu şirketlerle, çok kutuplu bir Nash denklemi içinde tıkanıp kalmış olsa da teknoloji bugün tek kutuplu bir dünyanın da muhtemelen dengeli bir Nash denklemi olabileceği kadar gelişmiş durumda. Mesela, dünyadaki herkesin aynı dil, kültür, değerler ve zenginlik seviyesini paylaştığı bir paralel evren düşünün. Bu evrende ulusların bir federasyondaki eyaletler gibi işlediği, ordularının olmadığı, yalnızca yasaları uygulayan polisin bulunduğu bir dünya hükümeti olduğunu hayal edin. Mevcut nüfusumuz böyle bir alternatif dengeye geçmeyi başaramasa ya da istekli olmasa bile şu anki teknoloji seviyemiz böylesine bir dünyayı koordine etmeye muhtemelen yetecektir.

Bu karışıma süper zeki YZ teknolojisi eklediğimizde kozmosumuzun hiyerarşik yapısına ne olur? Ulaştırma ve iletişim teknolojileri muhtemelen önemli ölçüde iyileşecek, böylece doğal olarak beklentimiz tarihsel trendin devam edeceği yönünde olacak. Daha da artan uzaklıklarda hiyerarşik seviyeler ortaya çıkacak, belki de nihai olarak güneş sistemlerini, galaksileri, süperkümeleri ve evrenimizin büyük kesimini de kapsayacak, bunu Bölüm 6'da inceleyeceğiz zaten. Aynı zamanda, yerelleşmenin en temel iticisi de devam edecek: Büyük mesafeler arasında gereksiz koordinasyon ziyandır. Stalin bile vatandaşlarının tuvalete ne zaman gittiklerini düzenlemeyi denemedi. Süper zeki YZ için, fizik kuralları ulaştırma ve iletişim teknolojisi için katı üst sınırlar koyarak gezegensel ve yerel ölçekte olan her şeyi çok yakından idare ve kontrol edebilecek ileri derece hiyerarşileri imkânsız kılacak. Andromeda galaksisinde bulunan süper zeki bir YZ, sizin gündelik kararlarınız için işe yarar tavsiyelerde bulunamayacak çünkü bunun için beş milyon yıldan fazla beklemeniz gerekecek (bu, ışık hızında iletişim kurabilmeniz için gidiş geliş geçecek zamandır). Aynı şekilde, dünyayı boydan boya geçip geri gelen

bir mesaj bu yolculuğu 0,1 saniyede gerçekleştirir (biz insanların düşündüğü zaman skalasında), bu yüzden de dünya büyüklüğünde bir YZ beyni yalnızca bir insan beyni kadar hızlı düşünebilir. Bir operasyonu saniyenin milyarda birinde gerçekleştirebilen küçük bir YZ için (ki bu süre bugünün bilgisayarları için normaldir), 0,1 saniye sizin için dört ay neyse odur. Bu yüzden de gezegen kontrol eden bir YZ tarafından yakından kontrol edilmesi, gündelik kararlarınızı Kolomb dönemi gemileriyle kıtalararası mektuplaşarak vermenize denk düşer.

Bilgi transferinin fizik kaynaklı hız sınırı da bu sebeple, bırakın evrenimizi, dünyamızı ele geçirmek isteyen herhangi bir YZ için bariz bir zorluk teşkil eder. Prometheus kaçmadan önce, zihin parçalanmasından sakınmak için çok düşünmüştü böylece dünyanın dört bir yanında çalışan YZ modüllerinin işbirliği yapmak ve tek bir birleşmiş yapı olarak hareket etmek için hedefleri ve teşvikleri mevcuttu. Prometheus'u kontrolde tutmaya çalışırken Omegalar'ın karşılaştığı sorunlar gibi, Prometheus da kendi kısımlarından hiçbirinin isyan etmeyeceğinden emin olmak istediğinde bir öz kontrol sorunu yaşadı. Halen daha bir YZ'nin ne büyüklükte bir sistemi doğrudan ya da bazı işbirlikçi hiyerarşiyle dolaylı olarak kontrol edebileceğini bilmiyoruz; hızlı kalkış ona belirgin bir stratejik avantaj vermiş olsa bile.

Özetle, süper zeki bir geleceğin nasıl kontrol edileceği sorusu baş döndürücü biçimde karmaşıktır ve açık biçimde, cevabı henüz bilmiyoruz. Bazıları her şeyin daha otoriter olacağını söylüyor; diğerleriyse büyük bireysel güçlenmeye yol açacağını iddia ediyor.

Cyborglar ve Yüklemeler

Bilimkurgunun başlıca konularından biri de insanların makinelerle birleşmesidir: Biyolojik bedenlerin teknolojik olarak zenginleşmiş

cyborglara dönüşmesiyle ya da zihinlerimizin makinelere yüklenmesiyle. Ekonomist Robin Hanson *The Age of Em* isimli kitabında yüklemelerle (*emülasyon/öykünüm* olarak da bilinen ve kısaca *Em* denen) kaynayan bir dünyada yaşamın nasıl olabileceğine dair baş döndürücü bir araştırma sunuyor. Bir yüklemeyi cyborg spektrumunun en ucu olarak düşünüyorum. Bu uçta, insandan geriye kalan tek şey yazılımdır. Hollywood cyborgları *Star Trek*'teki Borg gibi bariz biçimde mekanik olanlardan terminatörler gibi insanlardan ayrılamaz olan androidlere kadar uzanır. Kurgusal yüklemeler zekâda da bambaşka seviyelere sahiptir: *Black Mirror* dizisinin "White Christmas" bölümündeki insan seviyesi zekâdan *Transcendence* filmindeki süper insan zekâya kadar uzanır.

Eğer gerçekten de süper zekâ ortaya çıkarsa, cyborg ya da yükleme olma isteği güçlü olacaktır. 1988 yılında yayımlanan klasiği *Mind Children*'da Hans Moravec'in dediği gibi: "Giderek daha da inanılmaz olan keşiflerini bize anlayabileceğimiz bir bebek dili ile anlatmaya çalışırlarken ileri derece zeki makinelere uzun uzadıya aptalca bakarak harcamaksa kaderde olan, uzun yaşamak anlamını kaybeder." Gerçekten de teknolojik gelişmeye duyulan istek şimdiden bile o kadar çok ki damarlarında dolaşan tıbbi moleküllere ek olarak pek çok insanın gözlükleri, duyma aparatları, kalp pilleri ve protez uzuvları var. Bazı ergenler akıllı telefonlarına bağlı gibi yaşıyorlar ve eşim benimle laptopuma olan bağlılığım konusunda alay ediyor.

Günümüzün ileri gelen cyborg taraftarlarından biri de Ray Kurzweil. Kitabı *The Singularity is Near*'da bu yönelimin doğal devamının ilk olarak erken 2030'larda sindirim ve hormonal sistemimizin, kanımızın ve kalplerimizin yerini alacak, sonrasındaki yirmi yıl içerisinde ise iskeletimizin, derimizin, beynimizin ve vücutlarımızın kalan kısımlarını geliştirecek nanobotlar, zeki biyolojik geri bildirim sistemleri ve diğer teknolojiler olduğunu

savunuyor. İnsan vücutlarımızın estetik ve duygusal önemini hâlâ koruma eğiliminde olacağımızı ancak hem fiziksel hem de sanal gerçeklik olarak (yeni beyin-bilgisayar arayüzleri sayesinde) istenen her an görünüşlerini hızla değiştirebilecek şekilde onları yeniden tasarlayacağımızı tahmin ediyor. Moravec, cyborglaşmanın DNA'mızı iyileştirmenin çok ötesine gideceği konusunda Kurzweil'a katılıyor: "Genetik mühendisliği yapılmış bir süper insan ancak yalnızca DNA kılavuzluğunda gerçekleştirilen protein sentezince inşa edilebileceği handikabıyla tasarlanmış, ikinci sınıf bir robot olabilir." Dahası, insan vücudunu tamamen ortadan kaldırarak ve zihinlerimizi yükleyip tüm beyin emülasyonunu yazılımda yaparak daha da iyi bir konumda olabileceğimizi söylüyor. Böyle bir yükleme, sanal gerçeklikte ya da yürüme, uçma, yüzme, uzayda dolaşma ya da fizik kurallarınca izin verilen tüm diğer şeyleri yapabilecek bir robotun içinde bulunarak, ölüm ya da sınırlı bilişsel kaynaklar gibi gündelik dertlerle engellenmeden yaşayabilir.

Bu fikirler bilimkurgu gibi gelse de bilinen hiçbir fizik kanununu ihlal etmiyor, bu yüzden de en ilginç soru, olma *ihtimalleri* değil *olup olmayacakları* ve olacaksa da ne zaman olacakları. Bazı önemli düşünürler ilk insan seviyesi YGZ'nin bir yükleme olacağını ve süper zekâya doğru ilk adımın da böyle başlayacağını tahmin ediyorlar.*

Fakat bunun YZ araştırmacıları ve sinirbilimciler arasında azınlık görüşü olduğunu söylemek gerek. Çoğu, süper zekâya giden en hızlı yolun beyin emülasyonunu pas geçmek ve başka bir yol bulmak olduğunu düşünüyorlar; bu noktadan sonra beyin

* Bostrom'un açıkladığı gibi, saatlik ücretinden çok daha düşük bir maliyete önemli bir insan YZ araştırmacısının simülasyonunu yapma yetisi bir YZ şirketine iş gücünü önemli ölçüde artırma, ciddi oranda zenginlik biriktirme ve yinelemeli olarak daha iyi bilgisayarlar ve nihai olarak daha zeki zihinler inşa etmekteki gelişimlerini hızlandırma fırsatı tanır.

emülasyonuyla ilgilenebiliriz de ilgilenmeyebiliriz de. Sonuçta yeni bir teknolojiye giden en basit rota neden evrimin geldiği, kendini oluşturma, kendini tamir etme ve kendini kopyalama gibi gereksinimlerle sınırlanmış bir rota olsun ki? Evrim, besin kaynaklarının sınırlılığından ötürü, önemli ölçüde enerji verimliliği üzerine kurulmuştur; insan mühendisler tarafından kolayca anlaşılabilmek üzerine değil. Eşim Meia, havacılık endüstrisinin mekanik kuşlarla başlamadığını söylemeyi sever.[1] Gerçekten de mekanik kuşları nasıl yapacağımızı 2011 yılında, Wright kardeşlerin ilk uçuşundan yüz yıldan fazla süre geçtikten sonra bulduğumuzda, havacılık endüstrisi kanat çırpan mekanik kuş yolculuğuna karşı hiçbir ilgi göstermedi, hem de enerji verimliliği daha yüksek olmasına rağmen. Çünkü daha basit ve eski çözümümüz bizim yolculuk ihtiyaçlarımıza daha iyi uyuyordu.

Benzer bir şekilde, insan seviyesi düşünen makineler yapmanın evrimin bulduğu yoldan daha kolay yolları olabileceğini düşünüyorum. Eğer bir gün beyinlerimizi kopyalama ya da yüklemenin yolunu bulsak bile, ilk önce o basit çözümlerden birini keşfedeceğiz. Muhtemelen beyninizin kullandığı on iki vattan daha çok güç çekecek ama mühendisleri enerji verimliliği konusunda evrim kadar takıntılı olmayacak ve kısa süre içinde, zeki makinelerini daha verimli enerji kullananları tasarlamak için kullanacaklar.

Gerçekte Neler Olacak?

Kısa cevap elbette ki insanlık insan seviyesi YGZ inşa etmeyi başarırsa ne olacağına dair en ufak bir fikrimizin olmadığıdır. Bu sebeple, bu bölümü geniş spektrumda senaryoları incelemeye ayırdık. Oldukça kapsayıcı olmaya çalıştım, YZ araştırmacıları ve teknologların tartıştığını gördüğüm ya da duyduğum spekülasyonların her birini dâhil ettim: Hızlı kalkış/yavaş kalkış/kalkış olmama durumu, kontrolün insanlar/makineler/cyborglarda

olması, bir/çoklu güç merkezi *vs.* Bazı kişiler bana şunun ya da bunun asla olmayacağından emin olduklarını söylediler. Ancak bu aşamada alçak gönüllü olmanın ve ne kadar az şey bildiğimizi kabul etmenin akıllıca olacağını düşünüyorum çünkü yukarıda tartışılan her senaryo için, bu senaryoyu gerçek bir olasılık olarak gören en az bir değerli YZ araştırmacısı tanıyorum.

Zaman geçtikçe ve yolda belli başlı ayrımlara ulaştıkça, temel soruları cevaplamaya ve seçenekleri azaltmaya başlayacağız. İlk büyük soru, "İnsan seviyesi YGZ'yi yaratabilecek miyiz?" sorusudur. Bu bölümün önermesi yaratacağımızı iddia eder ancak asla olmayacağını, en azından yüzlerce yıl süreceğini düşünen YZ uzmanları da vardır. Buna zaman karar verecek! Daha önce de bahsettiğim gibi Porto Riko Konferansı'ndaki YZ uzmanlarının yaklaşık yarısı bunun 2055 yılına kadar olacağını tahmin ettiler. İki yıl sonrasında organize ettiğimiz devam konferansında bu tarih 2047'ye gelmişti.

İnsan seviyesi bir YGZ yaratılmadan önce, bu kilometre taşının ilk önce bilgisayar mühendisliği mi, zihin yükleme mi yoksa öngörülemeyen yeni bir yaklaşımla mı gerçekleşeceğine dair güçlü belirtiler almaya başlayabiliriz. Eğer şu anda alana hâkim olan YZ'ye bilgisayar mühendisliği yaklaşımı YGZ'yi yüzyıllar boyunca üretemezse, yüklemenin ilk olma şansı artacaktır. Tıpkı (pek de gerçekçi olmayan bir şekilde) *Transcendence* filminde olduğu gibi.

Eğer insan seviyesi YGZ daha olası hale gelirse, sonraki temel soruların cevapları hakkında daha mantıklı tahminler yapabileceğiz: "Hızlı kalkış mı olacak, yavaş kalkış mı yoksa kalkış olmayacak mı?" Yukarıda gördüğümüz gibi, hızlı kalkış dünyayı ele geçirmeyi kolaylaştırırken yavaş olanı pek çok rakibi olası kılıyor. Nick Bostrom bu soruyu sırasıyla YZ'yi daha zeki yapmak için gereken kaliteli çaba miktarı ve ilerleme katetme güçlüğü

anlamlarına gelen *optimizasyon gücü* ve *kontrol edilemezlik* hakkında yaptığı analizler üzerinden inceliyor. İlerlemenin ortalama hızı daha fazla optimizasyon gücü görevi yapmak için devreye alınırsa artıyor ve daha fazla kontrol edilemezlikle karşılaşıldığında düşüyor. YGZ insan seviyesine ulaşıp onu aşarken kontrol edilemezliğin artabileceğini de azalabileceğini de açıklayarak iki seçeneği de masada tutmanın garantici bir yaklaşım olacağını ileri sürüyor. Optimizasyon gücüne döndüğümüzde ise, YGZ'nin insan seviyesini aşması durumunda hızla artması büyük ihtimal. Bunun sebepleri de Omega senaryosunda gördüklerimizle aynı: Optimizasyonu ilerletmek için gereken ana girdi insanlardan değil makinenin kendinden geliyor, bu yüzden de makine ne kadar yetenekli hale gelirse o kadar hızlı gelişiyor (eğer kontrol edilemezlik sabit kalırsa).

Mevcut gücüne oranlı bir hızla artan gücü olan herhangi bir süreç için, sonuç gücünün düzenli aralıklarla iki katına çıkmasıdır. Böylesi bir büyümeye *katsal* deriz ve böylesi süreçlere *patlama* deriz. Eğer bebek yapma gücü nüfusun büyüklüğüne oranla artarsa, nüfus patlaması yaşarız. Eğer plütonyum füzyonu yapabilen nötronların oluşumu böyle nötronların sayısına oranla artarsa, nükleer patlama olur. Eğer makine zekâsı mevcut güce oranla bir hızda büyürse, zekâ patlaması yaşarız. Tüm böyle patlamalar güçlerini ikiye katlamaları için geçen zamanla nitelenir. Eğer bu zaman, Omega senaryosunda olduğu gibi, zekâ patlaması için saatler ve günler ise, elimizde hızlı bir kalkış olur.

Bu patlama zaman skalası, YZ'yi geliştirmenin yalnızca (birkaç saniye, dakika ya da saatte yaratılabilen) yeni yazılım mı yoksa (aylar ya da yıllar gerektiren) yeni donanım mı gerektirdiğine hayati olarak bağlıdır. Omega senaryosunda, Bostrom terimleriyle, ciddi bir *donanım fazlası* vardı: Omegalar, yüksek miktarda donanım sayesinde orijinal yazılımın düşük kalitesini telafi etmişlerdi.

Bu da Prometheus'un yalnızca yazılımını iyileştirerek defalarca kalitesini ikiye katlayabileceği anlamına geliyordu. İnternette bulunan verinin büyük bir kesimi ona yüklendiği için ciddi bir *içerik fazlası* da mevcuttu; Prometheus 1.0 çoğunu kullanabilecek kadar zeki değildi ancak Prometheus'un zekâsı geliştikçe, daha fazla öğrenmesi için gereken veri hiçbir gecikme olmadan *oradaydı.*

İnsan seviyesindeki işleri yapmanın maliyeti insan seviyesindeki saatlik ücretlerin altına düşmeden zekâ patlaması yaşayamayacağımız için YZ'yi çalıştırmanın donanım ve elektrik masrafı da hayati önem taşıyordu. Diyelim ki ilk insan seviyesi YGZ, ürettiği insan seviyesi işin saatlik maliyeti bir milyon dolar olacak şekilde Amazon bulutunda çalışabiliyor. Bu YZ büyük bir yenilik olacak ve kesinlikle manşetlerde kendine yer bulacaktır ancak yinelemeli bir öz gelişim sürecine girmeyecek çünkü onu geliştirmek için insanları kullanmayı sürdürmek hâlâ daha ucuz olacaktır. Diyelim ki bu insanlar kademe kademe maliyeti saatlik yüz bin dolar, on bin dolar, bin dolar, yüz dolar, on dolar ve en sonunda da bir dolara indirmeyi başardılar. Bilgisayarı kendini yeniden programlamak için kullanma maliyeti insan programcılara aynı iş için ödenenden çok daha az olduğunda, insanlar işten çıkarılabilir ve optimizasyon gücü bulut bilişim zamanı alınarak büyük oranda genişletilebilir. Bu daha da düşük maliyet anlamına gelir, daha fazla optimizasyon gücü sağlar ve zekâ patlaması başlar.

Bu da bizi son önemli sorumuza getiriyor: "Zekâ patlamasını ve sonrasını kim ya da ne kontrol edecek ve onların amaçları ne olacak?" Olası amaç ve sonuçları bir sonraki bölümde ve daha derinlemesine Bölüm 7'de inceleyeceğiz. Kontrol konusunu aydınlığa kavuşturmak için, hem YZ'nin ne kadar iyi kontrol edilebildiğini hem de YZ'nin ne kadar kontrol edebildiğini bilmek gerekiyor.

Nihai olarak ne olacağını düşünürsek, haritanın her köşesinde ciddi düşünürleri bulabilirsiniz. Bazıları sonucun kıyamet

olacağını iddia ederken diğerleri muazzam bir sonucun güvence altında olduğu konusunda ısrarcıdır. Bana göreyse, bu tuzak bir sorudur: “Ne olacak?” diye pasif biçimde sormak hatadır, sanki önceden kaderi belliymiş gibi! Eğer teknolojik olarak ileri bir uzaylı uygarlığı yarın buraya gelecek olsa, “Ne olacak?” diye düşünmek anlamlı olur çünkü uzay gemileri yaklaşmaktadır ve güçleri muhtemelen bizden çok daha ileridedir yani sonuç üzerinde hiçbir etkimiz yoktur. Eğer teknolojik olarak bizden üstün YZ destekli bir uygarlık onu biz inşa ettiğimiz için gelirse, biz insanlar sonuç üzerinde büyük etkiye sahip oluruz. YZ’yi oluştururken sarf ettiğimiz etkidir bu. Bu yüzden de sormamız gereken soru şudur: “Ne *olmalı*? Hangi geleceği istiyoruz?” Bir sonraki bölümde, YGZ’ye doğru yapılan mevcut yarışın olası sonuçlarının geniş spektrumunu inceleyeceğiz. Ben bu sonuçları iyiden kötüye nasıl sıralayacağınız konusunda meraktayım. Yalnızca nasıl bir gelecek istediğimiz konusunda ciddi ciddi düşündüğümüzde daha arzu edilen bir geleceğe doğru bir yol çizmeyi başarabileceğiz. Ne istediğimizi bilmiyorsak, ona ulaşmamız imkânsız olur.

SONUÇ:

- Eğer bir gün insan seviyesi YGZ'yi inşa etmeyi başarırsak, bu bizi çok geride bırakacak bir zekâ patlamasına yol açabilir.
- Eğer bir grup insan zekâ patlamasını kontrol ederse, yıllar içerisinde dünyayı ele geçirmeyi başarabilirler.
- Eğer insanlar zekâ patlamasını kontrol edemezlerse, YZ dünyayı daha hızlı biçimde kontrol altına alabilir.
- Hızlı bir zekâ patlamasının tekil bir dünya gücüne evrilmesi olasıyken, yıllar ya da on yıllar süren yavaş olanın yüksek sayıda görece bağımsız yapı arasında güç dengesiyle çok kutuplu bir senaryoya evrilmesi daha olasıdır.
- Yaşamın tarihi kendini işbirliği, rekabet ve kontrolle şekillenmiş karmaşık bir hiyerarşiyle organize etmiş olarak gösterir. Süper zekâ daha da büyük kozmik skalada işbirliğini mümkün kılmaya yatkındır ancak daha totaliter yukarıdan aşağı kontrole mi yoksa daha bireysel güçlenmeye mi yol açacağı belirsizdir.
- Cyborglar ve yüklemeler akla yatkındır ancak ileri makine zekâsına ulaşmak için en hızlı yol olmak zorunda değildirler.
- YZ'ye doğru mevcut yarışımızın zirvesi, bir sonraki bölümde inceleyeceğimiz mümkün sonuçların baş döndürücü spektrumuyla insanlığın başına gelmiş en iyi ya da en kötü şey olacak.
- Hangi sonucu tercih ettiğimiz ve bu yöne nasıl ilerleyeceğimiz hakkında ciddi ciddi düşünmeye başlamamız gerekir çünkü eğer ne istediğimizi bilmezsek ona ulaşmamız imkânsız olur.

Bölüm 5:



Sonuç: Sonraki 10000 Yıl

İnsan düşüncesinin ölümlü bir bedene bağlılığından kurtulduğunu düşünmek kolaydır, ölümden sonra yaşam düşüncesi yaygındır. Ancak bu olasılığı kabul etmek için mistik ya da dini bir duruş sergilemek gerekmez. Bilgisayarlar en ateşli mekanistler için bile bir model sunar.

Hans Moravec, *Mind Children*

Başta ben, yeni bilgisayar efendilerimizi selamlıyorum.

Ken Jennings, *Riziko!* oyununda IBM'in Watson'ına kaybettikten sonra

İnsanlar hamam böcekleri kadar alakasız olacaklar.

Marshall Brain

YGZ'ye doğru yarış başladı ve nasıl devam edeceğine dair hiçbir fikrimiz yok. Fakat bu bizi sonucun nasıl olmasını istediğimiz konusunda düşünmekten alıkoymamalı çünkü istediklerimiz sonucu etkileyecek. Kişisel olarak siz neyi tercih ediyorsunuz ve neden?

1. Süper zekânın olmasını ister misiniz?

2. İnsanların var olmaya devam etmesini mi cyborglarla değiştirilmesini mi yoksa yüklenmesini/simüle edilmesini mi istersiniz?
3. İnsanların mı makinelerin mi kontrolde olmasını istersiniz?
4. YZ'lerin bilinçli olmasını mı olmamasını mı istersiniz?
5. Pozitif deneyimleri maksimize, acıları minimize etmeyi mi istersiniz yoksa her şeyi oluruna bırakmayı mı?
6. Yaşamın kozmosa yayılmasını ister misiniz?
7. Bir uygarlığın sizin de desteklediğiniz büyük bir amaca doğru mücadele etmesini ister misiniz? Yoksa görüşlerini anlamsız ve banal bulsanız bile olumlu gözüken gelecek yaşam formlarını kabul eder misiniz?

Böyle bir düşünme ve diyaloğun gelişmesine yardımcı olmak için, Tablo 5.1'de özetlenen geniş senaryo yelpazesini inceleyelim. Tabii ki bu tüm seçenekleri sunmuyor ancak olasılık spektrumunu karşılayacak şekilde seçtim. Elbette kötü planlamadan dolayı yanlış bir sona varmak istemeyiz. Birden yediye sorulara verdiğiniz muhtemel cevapları not etmenizi ve bu bölümü okuduktan sonra fikrinizi değiştirip değiştirmediğinizi görmek için cevaplara yeniden bakmanızı öneririm! Bunu notları kıyaslayıp diğer okurlarla da tartışabileceğiniz http://AgeOfAi.org sitesinden de yapabilirsiniz.

YZ Sonuç Senaryoları	
Liberter ütopya	İnsanlar, cyborglar, yüklemeler ve süper zekâlar mülkiyet hakları sayesinde huzurlu şekilde bir arada yaşayabilirler.
İyiliksever diktatör	Herkes YZ'nin toplumu yönettiğini ve katı kurallar uyguladığını bilir ancak çoğu insan bunu iyi bir şey olarak görür.

Eşitlikçi ütopya	İnsanlar, cyborglar, yüklemeler ve süper zekâlar mülkiyetin ortadan kaldırılması ve garanti altına alınmış gelir sayesinde huzurlu şekilde bir arada yaşayabilirler.
Bekçi	Süper zekâ bir YZ başka bir süper zekânın yaratılmasını engelleyecek kadar az müdâhil olma hedefiyle yaratılır. Sonuç olarak, insan seviyesinin biraz altında zekâya sahip yardımcı robotlar bollaşır ve insan makine cyborglar da var olur ancak teknolojik ilerleme sonsuza dek engellenir.
Koruyucu tanrı	Özünde her şeyi bilen ve her şeye gücü yeten YZ, kendi kaderimizi kontrol ettiğimiz duygusunu koruyacağımız şekilde müdahale ederek insan mutluluğunu maksimize eder ve kendini o kadar iyi gizler ki bazı insanlar YZ'nin varlığından bile şüphe eder.
Esir edilmiş tanrı	Süper zekâ bir YZ insanlar tarafından hapsedilmiştir. Bu makine insan kontrolörlere bağlı olarak iyi ya da kötü yönde kullanılabilecek hayal dahi edilemeyen teknoloji ve zenginlik üretmek için kullanılır.
Fatihler	YZ kontrolü eline alır, insanların tehdit, dert ve kaynak ziyanı olduğuna karar verir ve bizim anlayamayacağımız bir yöntemle bizden kurtulur.
Nesil	YZ'ler insanların yerini alır ama nazikçe ayrılmamızı sağlarlar, onları kıymetli neslimiz olarak görürüz. Tıpkı kendilerinden daha zeki olan, onlardan öğrenen ama onların ancak hayal ettiği şeyleri gerçekleştiren çocuklarına karşı ebeveynlerin mutlu ve gururlu hissetmesi gibi – hem de bunu görmeye ömürleri yetmese bile.
Hayvan bakıcısı	Her şeye gücü yeten bir YZ, etrafında bazı insanları tutar. Bu insanlar kendilerini hayvanat bahçesindeki hayvanlar gibi görür ve kaderlerine hayıflanırlar.
1984	Süper zekâya doğru teknolojik ilerleme, bir YZ değil ama bazı YZ araştırmalarının yasaklandığı insan yönetimli Orwellci bir gözetim toplumu tarafından engellenir.
Geçmişe intikal	Süper zekâya doğru teknolojik ilerleme, Amish tarzı teknoloji öncesi bir topluma dönülerek önlenir.
Özyıkım	Süper zekâ asla yaratılamaz çünkü insanlık kendini başka yollarla ortadan kaldırır (mesela nükleer ya da iklim krizinin ortaya çıkardığı biyoteknolojik bir felaket).

Tablo 5.1: YZ Sonuç Senaryoları Özeti

Senaryo	Süper zekâ mevcut mu?	İnsanlar mevcut mu?	İnsanlar kontrolde mi?	İnsanlar güvende mi?	İnsanlar mutlu mu?	Bilinç mevcut mu?
Liberter ütopya	Evet	Evet	Hayır	Hayır	Karışık	Evet
İyiliksever diktatör	Evet	Evet	Hayır	Evet	Karışık	Evet
Eşitlikçi ütopya	Hayır	Evet	Evet?	Evet	Evet?	Evet
Bekçi	Evet	Evet	Kısmen	Potansiyel olarak	Karışık	Evet
Koruyucu tanrı	Evet	Evet	Kısmen	Potansiyel olarak	Karışık	Evet
Esir edilmiş tanrı	Evet	Evet	Evet	Potansiyel olarak	Karışık	Evet
Fatihler	Evet	Hayır	-	-	-	?
Nesil	Evet	Hayır	-	-	-	?
Hayvan bakıcısı	Evet	Evet	Hayır	Evet	Hayır	Evet
1984	Hayır	Evet	Evet	Potansiyel olarak	Karışık	Evet
Geçmişe intikal	Hayır	Evet	Evet	Hayır	Karışık	Evet
Özyıkım	Hayır	Hayır	-	-	-	Hayır

Tablo 5.2: YZ Sonuç Senaryolarının Özellikleri

Liberter Ütopya

İnsanların barış içerisinde teknolojiyle birlikte ve bazı durumlarda da onunla birleşmiş bir şekilde yaşadığı, pek çok fütürist ve bilimkurgu yazarının tasavvurundaki bir senaryoyla başlayalım:

Dünyadaki (ve ötesindeki – bundan sonraki bölümde daha çok bahsedeceğiz) yaşam her zamankinden daha çoğulcudur. Dünyanın uydu görüntülerine bakarsanız, makine bölgeleri, karışık bölgeler ve insan bölgelerini kolayca ayırt edebilirsiniz. Makine bölgeleri biyolojik yaşamın olmadığı, her atomun en verimli şekilde işe koşulması hedeflenen robot kontrollü devasa fabrikalar ve işleme tesisleridir. Makine bölgeleri dışarıdan monoton ve

sönük gözükse de içeride, sanal dünyalarda inanılmaz deneyimler yaşanır ve devasa işlemler evrenimizin sırlarını çözüp dönüştürücü teknolojiler geliştirirken, baş döndürücü bir canlılık vardır. Dünya rekabet içerisinde olan ya da işbirliği yapan pek çok süper zeki zihne sahiptir ve her biri de makine bölgelerinin içindedir.

Karışık bölgelerin sakinleri bilgisayarlar, robotlar, insanların kendine has ve çılgın bileşenleri ve tüm bu üçünün melezlerinden oluşur. Hans Moravec ve Ray Kurzweil gibi fütüristlerin hayal ettiği üzere, insanların büyük bir çoğunluğu bedenlerini belli aşamalarda cyborglara teknolojik olarak evriltmiştir. Bazıları zihinlerini yeni donanımlara yükleyerek insan ve makine arasındaki ayrımı bulandırmıştır. En zeki varlıklar sabit bir fiziksel formdan yoksundur. Onun yerine, bilgisayarlar arasında sürekli hareket edebilen bir yazılım olarak var olurlar ve fiziksel dünyada kendilerini robotik bedenler üzerinden ifade ederler. Bu zihinler kendilerini kopyalayıp birleşebildiği için, "nüfus büyüklüğü" devamlı değişir. Fiziksel özlerinden bağımsız olmak yaşama karşı çok farklı bir bakışa sahip olmalarına sebep olur. Çok daha az bireysel hissederler çünkü çok da önemli olmayan bir şekilde diğerleriyle bilgi ve deneyim modüllerini paylaşabilirler ve öznel olarak ölümsüz hissederler çünkü zaten kendilerinin kopyalarını yapabiliyorlardır. Bir bakıma, yaşamın temel yapıları zihinler değil deneyimlerdir. Sıra dışı ve inanılmaz deneyimler yaşanır çünkü devamlı kopyalanır ve başka zihinlerce yeniden tadılır, bu esnada sıkıcı deneyimler ise daha iyileri için alan açmak amacıyla sahipleri tarafından silinir.

Etkileşimlerin birçoğu hız ve kolaylık sebebiyle sanal ortamlarda gerçekleşse de pek çok zihin fiziksel bedenlerini kullanarak etkileşimde bulunmaktan ve aktivitelere dâhil olmaktan hoşlanır. Mesela Hans Moravec, Ray Kurzweil ve Larry Page'in yüklenmiş versiyonları sırayla sanal gerçeklikler yaratıp onu beraber keşfet-

mek gibi bir geleneğe sahiptirler ama arada bir de olsa kendilerini uçan kanatlı robotlara aktarıp gerçek dünyada da uçmanın tadını çıkarırlar. Karışık bölgelerin sokaklarında, göklerde ve göllerinde dolaşan bazı robotlar da aynı şekilde yüklenmiş ve artırılmış insanlar tarafından kontrol edilir. Bu insanlar kendilerini karışık bölgelerde vücuda sokmuşlardır çünkü insanların ve kendilerine benzeyenlerin arasında olmak onlara mutluluk verir.

Yalnızca insanların olduğu bölgelerde ise insan seviyesi genel zekâ ya da üzeri zekâya sahip makineler ve teknolojik olarak zenginleştirilmiş biyolojik organizmalar yasaklanmıştır. Burada, yaşam bugünden çok da farklı değildir. Tek farkı daha zengin ve kolay olmasıdır: Fakirlik genel anlamda ortadan kaldırılmıştır ve bugünün hastalıklarının çoğunun tedavisi bulunmuştur. Bu bölgelerde kalmak isteyen küçük bir grup insan diğer herkese göre daha sınırlı ve düşük bir farkındalık düzeyinde yer alırlar ve diğer bölgelerde daha zeki zihinlerin neler yaptığına dair de kısıtlı bir anlayışa sahiptirler. Ancak pek çoğu hayatlarından memnundur.

YZ Ekonomisi

Tüm işlemlerin büyük bir kısmı orada yaşayan ve rekabet içerisinde olan süper zeki YZ'ler tarafından sahiplenilen makine bölgelerinde gerçekleşir. Üstün zekâlarının ve teknolojilerinin özelliği olarak, başka hiçbir yapı onların güçlerine kafa tutamaz. Bu YZ'ler de özel mülkiyetin korunması haricinde hiçbir kuralı olmayan liberter bir yönetim sisteminde işbirliği yapmayı ve koordine olmayı kabul etmişlerdir. Bu mülkiyet hakları, insanlar da dâhil olmak üzere tüm zeki varlıklara genişletilmiştir ve yalnızca insanların bulunduğu bölgelerin nasıl var olabildiğini açıklar. Öncesinde, insan grupları bir araya gelmiş ve bölgelerinde insan olmayanlara mülk satmanın yasak olacağına karar vermişlerdi.

Teknolojilerinden dolayı, süper zeki YZ'ler insanlara göre, Bill Gates'in evsiz bir dilenciye oranla zengin olduğundan çok daha yüksek oranda zengin olmuşlardır. Ancak sadece insanların bulunduğu bölgelerdeki kişiler yine de bugünkü pek çok insandan daha zengindirler. Ekonomileri makinelerinkinden ayrışmıştır, bu yüzden de makinelerin başka yerlerdeki varlığının onların üzerinde pek az etkisi vardır. Bunun bir istisnası anlayıp kendileri için üretebilecekleri faydalı teknolojilerdir; bugün Amish ve çeşitli teknolojiden feragat etmiş yerli kabilelerin standartlarının en az eski günlerdeki kadar iyi olması gibi. İnsanların makinelerin ihtiyacı olan hiçbir şeyi satmıyor olması önemli değildir çünkü makinelerin karşılığında bir şeye ihtiyacı yoktur.

Karışık bölgelerde, YZ'ler ve insanlar arasındaki zenginlik farkı daha gözle görülür durumdadır. Bu da arazinin (makinelerin almak isteyeceği, insanlar tarafından sahiplenilen tek ürün) diğer ürünlere kıyasla astronomik rakamlarda satılmasına yol açar. Toprak sahibi olan çoğu insan da arazilerinin küçük bir kesimini YZ'lere kendilerini ve gelecekteki nesillerini yüklemeleri için garanti edilmiş vatandaşlık geliri karşısında satar. Bu onları çalışma zahmetinden kurtarır ve hem fiziksel hem de sanal gerçeklikte, ucuz makine üretimi malların ve hizmetlerin inanılmaz bolluğunun tadını çıkarabilmeleri için boş vakit sağlar. Makineler için ise, karışık bölgeler işten çok eğlence alanlarıdır.

Bu Neden Asla Gerçekleşmeyebilir?

Cyborglar ya da yüklemeler olarak yaşayacağımız maceralar konusunda çok heyecanlanmadan önce, bu senaryonun neden asla gerçekleşmeyebileceğine dair bazı sebepleri inceleyelim. İlk olarak, zenginleştirilmiş insanlar (cyborglar ve yüklemeler) olarak yaşamamızın iki olası yolu var:

1. Bunu yapmayı kendimiz öğreniriz.
2. Bizim için bunu öğrenecek süper zeki makineler üretiriz.

Eğer ilk önce birinci yol gerçekleşirse, doğal olarak cyborglar ve yüklemelerle kaynayan bir dünya oluşabilir. Fakat önceki bölümde tartıştığımız üzere, çoğu YZ araştırmacısı tersinin daha olası olduğunu söylüyor çünkü zenginleştirilmiş ya da dijital beyinler yapmak, sıfırdan süper insan YGZ üretmekten daha zor. Tıpkı mekanik kuşları üretmenin uçak yapmaktan daha zor olması gibi. Güçlü makine YZ'si üretildikten sonra, cyborg ya da yüklemelerin yapılabileceği de kesin değildir. Eğer Neandertaller yüz bin yıl daha yaşayıp evrilebilseler ve zekileşselerdi, her şey onlar için daha iyi olabilirdi ancak Homo Sapiens onlara bu vakti tanımadı.

İkincisi, cyborglar ve yüklemelerle dolu bu senaryo gerçekleşse bile, dengeli olacağı ve devam edeceği açık değildir. Neden çoklu süper zekâlar arasındaki güç dengesi bin yıllar boyunca devam etsin, yerini YZ'lerin birleşmesine ya da en zekinin hepsini ele geçirmesine bırakmasın ki? Dahası, makineler neden insanların mülkiyet haklarına saygı duyup onları etrafta tutsun, sonuçta hiçbir şey için insanlara ihtiyaçları yok ve her işi insanlardan daha iyi ve ucuza kendileri yapabiliyorlar? Ray Kurzweil doğal ve zenginleştirilmiş insanların yok olmaktan korunacağını çünkü "insanların makineleri yarattıkları için YZ'lerden saygı göreceğini" söylüyor.[1] Ancak, Bölüm 7'de konuşacağımız üzere, YZ'lere insani vasıflar atfetme yanılgısına düşmemeli ve insan benzeri minnet duyguları olduğunu düşünmemeliyiz. Gerçekten de biz insanlar minnet eğilimiyle dolu olsak da doğum kontrolü kullanarak kendi entelektüel yaratıcımızın (DNA'mızın) amaçlarına yeterince saygı göstermeyebiliyoruz.

YZ'lerin insan mülkiyet haklarına saygı duyacağı varsayımını kabul etsek bile, arazimizin büyük bir kısmını başka yollarla ele geçirebilir, son bölümde incelediğimiz süper zeki ikna güçleriyle insanları ikna edip lüks yaşam için arazilerini satmaya ikna edebilirler. Yalnızca insanların yaşadığı kesimlerde, insanları arazi satışları için politik kampanyalar düzenleme konusunda ayartabilirler. Sonuçta, en sıkı biyo-Ludditelar bile hasta bir çocuğun hayatını kurtarmak ya da ölümsüzlük kazanmak için arazisinin bir kısmını satmak isteyebilir. Eğer insanlar eğitimli, eğlenmiş ve meşgulse, düşen doğum oranları nüfusu makinelerin müdahalesi olmadan azaltır, tıpkı şu an Almanya ve Japonya'da olduğu gibi. Bu da birkaç bin yıl içerisinde insanların yok olmasına yol açabilir.

Dezavantajlar

En sıkı savunucuları için, cyborglar ve yüklemeler herkes için tekno-mutluluk ve uzun yaşam vaadi sunuyor. Gerçekten de gelecekte yüklenebilecek olma ihtimali yüzden fazla insanın öldükten sonra Arizona merkezli Alcor firmasında beyinlerini dondurma konusunda motive etmişti. Eğer bu teknoloji gelirse, bunun herkes için uygulanabilir olacağı kesin değildir. En zenginlerin büyük kısmı bunu kullanacaktır ama başka kime piyango vuracak? Teknoloji ucuzlasa bile, çizgiyi nereden çekeceğiz? Ciddi beyin hasarı olanlar da yüklenecek mi? Her gorili yükleyecek miyiz? Her karıncayı? Her bitkiyi? Her bakteriyi? Geleceğin uygarlığı obsesif kompulsif stokçular gibi davranıp her şeyi yüklemeyi mi deneyecek yoksa Nuh'un Gemisi gibi her türden birkaç ilginç örnek mi yüklenecek? Yoksa her tür insandan birkaç temsilci mi olacak? O zaman mevcut olacak geniş ölçüde zeki yapılar için yüklenmiş bir insan simüle edilmiş bir fare ya da salyangozun şu an bize ilginç geldiği kadar ilginç gelir ancak. 1980'lerin eski

tablolama programlarını bir DOS emülatöründe yeniden canlandırma teknolojisine sahip olsak bile, hiçbirimiz bunu yapmakla ilgilenmiyor.

Pek çok insan bu liberter ütopya senaryosunu beğenmeyebilir çünkü önlenebilecek acıyı mümkün kılıyor. Tek kutsal ilke mülkiyet hakkı olduğu için, bugünün dünyasında bolca bulunan acının insan ve karışık bölgelerde devam etmesini hiçbir şey engellemiyor. Bazı insanlar zengin olurken, diğerleri sefalet içinde veya sözleşmeli köle olarak yaşayabilir ya da şiddet, korku, baskı veya depresyondan muzdarip olabilir. Örneğin, Marshall Brain'in 2003 yılında yayımlanan romanı *Manna* liberter ekonomik sistemde YZ gelişiminin Amerikalıların çoğunu nasıl işsiz bırakacağını ve yaşamlarının kalanını kasvetli ve donuk robot işletmeli sosyal refah konut projelerinde geçirmek zorunda kalacaklarını anlatıyordu. Tıpkı çiftlik hayvanları gibi zenginlerin onları asla görmek zorunda kalmayacağı kısıtlı bir durumda beslenecek, sağlıklı ve güvende tutulacaklardı. Suda bulunan doğum kontrol ilaçları çocuklarının olmamasını garanti altına alacak, böylece nüfus kademe kademe azalarak kalan zenginlerin robot üretimi zenginlikten daha büyük paylar almasını sağlayacaktı.

Liberter ütopya senaryosunda, acı insanlarla sınırlı değildir. Eğer bazı makineler bilinçli duygusal deneyimlere sahip olurlarsa, onlar da acı çekebilir. Mesela, kinci bir psikopat yasal bir şekilde düşmanının bir kopyasını yükleyebilir ve onu sanal dünyada korkunç işkencelere tabi tutarak gerçek dünyada biyolojik olarak mümkün olamayacak yoğunluk ve uzunlukta acı verebilir.

İyiliksever Diktatör

Şimdi de tüm bu acı türlerinin olmadığı çünkü tek bir iyiliksever süper zekânın dünyayı yönettiği ve insan mutluluğu modelini üst düzeye çıkarmak üzere tasarlanmış katı kuralları uyguladığı bir senaryoyu inceleyelim. Önceki bölümde, insan toplumunun

gelişmesini istemesini sağlamanın bir yolunu bulduktan sonra kontrolü Prometheus'a bıraktıkları ilk Omega senaryosunun olası sonuçlarından biri de budur.

Diktatör YZ tarafından geliştirilmiş inanılmaz teknolojiler sayesinde, insanlık fakirlik, hastalık ve diğer düşük teknoloji problemlerinden kurtulur ve tüm insanlar lüks bir eğlence hayatı yaşar. Tüm temel ihtiyaçları karşılanırken, YZ kontrolündeki makineler tüm gerekli mal ve hizmetleri üretir. Suç ortadan kalkmıştır çünkü YZ özünde her şeyi bilir ve kurallara uymayan bir kimseyi etkili biçimde cezalandırabilir. Herkes bir önceki bölümde anlatılan güvenlik bileziğini (ya da onun daha basit implant versiyonunu) takar; bu bilezik, gerçek zamanlı gözetim, cezalandırma, yatıştırma ve idamı gerçekleştirebilir. Herkes ileri derece gözetim ve polis faaliyeti içerisinde bir YZ diktatörlüğünde yaşadığını bilir ancak çoğu insan bunu iyi bir şey olarak görür.

Süper zeki YZ diktatör, genlerimizde kodlanmış evrimleşmiş tercihlerimize göre insan ütopyasının nasıl olduğunu bulmak ve bunu uygulamak hedefine sahiptir. YZ'yi ortaya çıkaran insanların zekice öngörüsüyle, herkese damardan morfin vermek suretiyle, insanların özbildirimine göre mutluluğunu artırmayı denemez. Bunun yerine YZ, insan gelişimine dair oldukça incelikli ve karmaşık bir tanım kullanır ve dünyayı insanların içinde yaşamaktan keyif aldığı hayli zenginleştirilmiş bir hayvanat bahçesine çevirir. Sonuç olarak, pek çok insan yaşamlarını oldukça doyurucu ve anlamlı bulur.

Sektör Sistemi

Çoğulculuğa değer veren ve farklı insanların farklı tercihleri olduğunu anlayan YZ, dünyayı insanların seçmesi ve kafa dengi insanlarla bir arada olmaları için farklı sektörlere ayırır. İşte bazı örnekler:

- Bilgi sektörü: Burada YZ, seçtiğiniz herhangi bir konuda istediğiniz her şeyi öğrenmenizi sağlayacak, derin sanal gerçeklik deneyimleri sunan optimize bir eğitim sağlar. İsteğe bağlı olarak, bazı fikirlerin size hemen açıklanmamasını ama onlara doğru yönlendirilmeyi seçebilir ve kendiniz keşfetmenin tadını çıkarabilirsiniz.
- Sanat sektörü: Burada müzik, sanat, edebiyat ve diğer yaratıcı ifade formlarının tadını çıkarabilir, onları üretebilir ve paylaşabilirsiniz.
- Hedonistik (hazcı) sektör: Buraya parti sektörü de denir ve leziz yemekler, tutku, mahremiyet ya da çılgın eğlence için buradan daha iyisi yoktur.
- Dinî sektör: Kurallarının katı biçimde uygulandığı farklı dinlere tekabül eden birçok sektör vardır.
- Doğal yaşam sektörü: Güzel plajlar, hoş göller, büyüleyici dağlar ya da fantastik fiyortlar arıyorsanız, hepsi burada.
- Geleneksel sektör: Burada kendi yemeğinizi yetiştirebilir ve geçmişteki gibi topraktan geçinebilirsiniz ama hastalıklar ya da kıtlıktan endişe etmenize gerek yok.
- Oyun sektörü: Eğer bilgisayar oyunlarını seviyorsanız, YZ sizin için akıllara durgunluk verecek seçeneklere sahip.
- Sanal sektör: Fiziksel bedeninizden uzakta bir tatil istiyorsanız, YZ onu sular, besler, egzersizlerini yapar ve temizlerken siz sinir implantlarıyla sanal dünyaları deneyimlersiniz.
- Hapis sektörü: Eğer kuralları çiğnerseniz, yeniden eğitim için buraya gelirsiniz, elbette anında idam edilmediyseniz.

Bu "geleneksel" temalı sektörlere ek olarak, bugünün insanlarının anlayamayacağı modern temalara sahip başkaları da vardır. İnsanlar farklı sektörler arasında istediklerine geçiş yapmakta

serbesttirler, bu da YZ'nin sesten hızlı ulaşım sistemi sayesinde çok az zaman alır. Mesela, bilgi sektöründe yoğun bir hafta geçirip YZ'nin keşfettiği tüm fizik yasalarını öğrendikten sonra, hafta sonu ipinizi koparıp hedonistik sektörde takılmayı, sonra da birkaç günlüğüne doğal yaşam sektöründeki sahil tesisinde dinlenmeyi isteyebilirsiniz.

YZ iki aşama kural uygular: evrensel ve yerel. Evrensel kurallar her sektörde geçerlidir, mesela diğer insanlara zarar verme, silah yapma ya da rakip bir süper zekâ inşa etmeyi denemenin yasak olması gibi. Her sektörün ek yerel kuralları vardır, bunlar da çeşitli ahlaki kuralları içerir. Sektör sistemi de birbirine uymayan değerlerle uğraşmaya yardımcı olur. En çok yerel kural hapis sektöründe ve bazı dinî sektörlerdedir. Sakinlerinin hiçbir yerel kural olmamasıyla gurur duyduğu liberter sektör de vardır. Tüm cezalar, yerel olanları bile, YZ tarafından icra edilir çünkü bir insanın başka bir insana ceza vermesi evrensel zarar vermeme kuralını ihlal eder. Yerel bir kuralı ihlal ederseniz YZ size seçenek sunar (eğer hapis sektöründe değilseniz): Ya cezayı kabul edeceksiniz ya da o sektörden sonsuza dek atılacaksınız. Mesela, iki kadın homoseksüelliğin hapisle cezalandırıldığı bir sektörde (bugün birçok ülkede olduğu gibi) romantik bir ilişki yaşarsa, YZ onlara hapishaneye gitmek ile o sektörü tamamen terk etmek, eski arkadaşlarıyla bir daha asla görüşmemek (onlar da ayrılmazlarsa tabii) gibi iki seçenek sunar.

Hangi sektörde doğmuş olurlarsa olsunlar, tüm çocuklar insanlık hakkında bütünlüklü bir bilgiyi ve isterlerse diğer sektörleri ziyaret edip oralara taşınmakta özgür olduklarını da içeren temel bir eğitimi YZ'den alırlar.

YZ yüksek sayıda farklı sektörü kısmen bugün mevcut olan insan çoğulluğuna değer vermek üzere üretildiği için tasarlamıştır. Ancak her sektör bugünün teknolojisinin izin vereceğinden daha mutludur çünkü YZ açlık ve suç gibi tüm geleneksel sorunları

ortadan kaldırmıştır. Mesela hedonistik sektördeki insanların cinsel yolla bulaşan hastalıklardan (çünkü hepsi ortadan kalkmıştır), içki sersemliğinden ya da bağımlılıktan (YZ hiçbir negatif yan etkisi olmayan mükemmel uyuşturucular geliştirmiştir) endişe etmelerine gerek kalmamıştır. Gerçekten de hiçbir sektörde hiç kimse hastalıklardan endişe etmez çünkü YZ insan vücudunu nanoteknolojiyle tedavi edebilmektedir. Çoğu sektörün sakini tipik bilimkurgu vizyonunu sönük bırakacak yüksek teknoloji mimariden de faydalanabilir.

Özet olarak, liberter ütopya ve iyiliksever diktatör senaryolarının ikisi de ileri derece YZ temelli teknoloji ve zenginliği içerirken, kimin kontrolde olacağı ve hedefleri açısında farklılaşır. Liberter ütopyada, teknoloji ve mülkiyete sahip olanlar bununla ne yapacaklarına karar verirler, bu senaryoda ise diktatör YZ'nin sınırsız gücü vardır ve nihai hedefi belirler: dünyayı herkesi içerecek şekilde ve herkesin kişisel isteğine göre temalandırılmış bir zevk turuna çevirmek. YZ insanların pek çok alternatif mutluluk yolu arasından seçim yapmasına izin verdiği ve maddi ihtiyaçlarını karşıladığı için, birisi acı çekerse bu o kişinin seçimi olacaktır.

Dezavantajlar

İyiliksever diktatörlük pozitif enerjiyle dolup taşsa ve acıdan uzak olsa da pek çok insan her şeyin daha iyi olabileceğini düşünebilir. İlk olarak, bazı insanlar insanlığın kendi toplumlarını ve kaderlerini şekillendirmede daha fazla özgürlüğe sahip olmalarını dileyebilir ancak bu dileklerini kendilerine saklarlar çünkü herkesi yöneten makinenin büyük gücüne karşı gelmenin intihara eşdeğer olduğunu bilirler. Bazı gruplar istedikleri kadar çocuk yapma özgürlüğü isteyebilir ve YZ'nin nüfus kontrolüyle sürdürülebilirliği sağlamasına gücenebilirler. Silah severler silah yapma ve kullanma yasağından nefret edebilirler ve bazı bilim insanları

kendi süper zekâlarını yapma yasağından hoşlanmayabilirler. Pek çok insan diğer sektörlerde olan şeylere karşı ahlaki bir öfke besleyebilir, çocuklarının oraya taşınmasından endişe duyabilir ve her yerde kendi ahlaki yasalarını empoze etme özgürlüğü için yanıp tutuşabilirler.

Zamanla daha çok insan YZ'nin onlara istedikleri her deneyimi verdiği sektörlere taşınmak isteyebilir. Hak ettiğiniz şeyi yaşadığınız geleneksel cennet vizyonlarının tersine, bu Julian Barnes'ın 1989'da yayımlanan *On Buçuk Bölümde Dünya Tarihi* isimli romanındaki (ya da 1960 yılında yayımlanan *Alacakaranlık Kuşağı* bölümü "A Nice Place to Visit") gibi herkesin arzu ettiği şeyi aldığı "yeni cennet" kavramıdır. Paradoksal olarak, pek çok insan istedikleri şeyi elde etmekten dolayı hayıflanırlar. Barnes'ın hikâyesinde ana karakter oburluk ve golften ünlülerle sekse kadar çeşitli arzularına dalarak uzun zaman geçirir ancak sonunda usanca yenik düşer ve yok edilmeyi ister. İyiliksever diktatörlükteki pek çok insan da benzeri bir kaderle karşı karşıya gelir, tatminkâr yaşamlardır bunlar ama nihai olarak anlamsızdırlar da. Bilimsel yeniden keşiften kaya tırmanışına dek, insanlar suni zorluklar yaratabiliyor olsa da herkes gerçekten bir zorluk olmadığını, bunun bir eğlence olduğunu bilir. İnsanların bilim yapmak ya da başka şeyleri keşfetmekle uğraşmasının hiçbir gerçek anlamı yoktur çünkü YZ çoktan halletmiştir bunları. İnsanların hayatlarını iyileştirmek için bir şey üretmeye çalışmasında da bir anlam yoktur çünkü hepsini YZ'den isteyerek elde edebilirler.

Eşitlikçi Ütopya

Meşakkat barındırmayan diktatörlüğe karşı olarak, şimdi de süper zeki YZ'nin olmadığı ve insanların kendi kaderlerinin efendisi olduğu bir senaryoyu inceleyelim. Bu Marshall Brain'in 2003'te yayımlanan romanı *Manna*'da geçen "dördüncü nesil uygarlık-

tır." İnsanların, cyborglar ve yüklemelerin mülkiyet haklarından dolayı değil, mülkiyet ortadan kaldırıldığı ve gelir garanti altına alındığı için barış içinde bir arada yaşadıkları liberter ütopyanın ekonomik antitezi bir yaklaşımdır bu.

Mülkiyetsiz Yaşam

Temel fikir açık kaynak yazılım hareketinden gelmektedir: Eğer yazılımı kopyalamak bedavaysa, herkes bunu ihtiyacı olduğu kadar kullanabilir ve sahiplik ile mülkiyet tartışmaları anlamsız olur.* Arz ve talep yasalarına göre, maliyet az bulunurluğu yansıtır. Bu yüzden de eğer arz özünde sınırsız olursa fiyat ihmal edilebilir hale gelir. Bu mantıkla, tüm fikrî mülkiyet hakları ortadan kaldırılır. Hiçbir patent, telif hakkı ya da ticari markalı tasarım yoktur; insanlar iyi fikirlerini paylaşır ve herkes bunları kullanmakta özgürdür.

Gelişmiş robotlar sayesinde bu mülkiyetsizlik fikri yalnızca yazılım, kitaplar, filmler ve tasarımlar gibi bilgi ürünlerine değil, evler, arabalar, kıyafet ve bilgisayarlar gibi maddi ürünlere de uygulanabilir. Tüm bu ürünler de sonuçta farklı şekillerde düzenlenmiş atomlardır ve atom eksiğimiz olmadığı için bir insan herhangi bir ürünü istediği bir robot ağını ücretsiz bir şekilde ve müsait bir açık kaynak tasarımı kullanarak üretebilir. Kolayca geri dönüştürülebilir materyaller daha çok kullanılır ki böylece bir kişi kullandığı bir nesneden sıkıldığında, robotlar atomlarını başka birinin istediği başka bir şey olarak yeniden düzenleyebilsin. Bu şekilde, tüm kaynaklar geri dönüşüme tabidir ve hiçbir şey tamamen yok edilmez. Bu robotlar yeterince yenilenebilir enerji

* Bu fikir, Aziz Augustin'e kadar geri gitmektedir: "Eğer bir şey başkalarıyla paylaşmaktan dolayı azalmıyorsa, yalnızca sahiplenilip paylaşılmıyorsa hakça sahiplenilmemiştir."

üreten tesis (rüzgâr, güneş vb.) inşa edip kullandıkları için enerji de temelde ücretsizdir.

Takıntılı biriktiricilerin sürekli yeni ürün ya da arazi talep ederek başkalarını muhtaç bırakmalarını engellemek için her insan hükümetten bir vatandaşlık geliri alarak bu parayı istedikleri ürünlerde ya da yaşayacak yerleri kiralamakta kullanabilir. Kimsenin daha fazla para kazanmayı denemesi için hiçbir teşvik yoktur çünkü temel gelir her türlü anlamlı ihtiyacı karşılayabilecek kadar yüksektir. Denemek de boşuna olacaktır çünkü entelektüel ürünleri bedavaya veren insanlar ve bedavaya materyal ürünler üreten robotlarla rekabet etmek zorundadırlar.

Yaratıcılık ve Teknoloji

Fikrî mülkiyet hakları bazen yaratıcılık ve keşfin anası olarak gösterilir. Ancak Marshall Brain insan yaratıcılığının pek çok iyi örneğinin –bilimsel keşiflerden edebiyat, sanat, müzik ve tasarımdaki üretime– kâr amacıyla değil, merak, yaratma isteği ya da çevresinden alacağı takdir gibi insani duygularla gerçekleştiğini ileri sürüyor. Para Einstein'ın özel görelilik teorisini bulmasında, Linus Torvalds'ın ücretsiz Linux işletim sistemini yapmasından daha fazla bir motivasyon kaynağı sağlamamıştı. Tam aksine, pek çok insan bugün, para kazanmak için daha az yaratıcı aktiviteye zaman ve enerji ayırmak zorunda oldukları için tüm yaratıcı potansiyellerini gerçekleştirmeyi başaramıyor. Bilim insanlarını, sanatçıları, mucitleri ve tasarımcıları gündelik işlerden kurtaran ve onlara arzuları doğrultusunda üretebilme fırsatı sunan Marshall Brain'in ütopya toplumu, günümüzden daha yüksek seviye yenilik yaşayacak ve aynı şekilde üstün teknoloji ve yaşam standardına sahip olacaktır.

İnsanların geliştirdiği böyle bir yeni teknoloji de Vertebrane adı verilen bir tür hiper-internettir. İsteyen tüm insanları kablosuz biçimde sinir implantlarıyla birbirine bağlayarak yalnızca düşünceyle dünyanın tüm ücretsiz bilgisine anında zihin erişimi sağlar. Paylaşmak istediğiniz herhangi bir deneyimi yükleyerek başkalarınca yeniden deneyimlenmesini sağlayabilir ve seçtiğiniz sanal deneyimleri indirerek algılarınıza gelen deneyimi değiştirmenize izin verebilir. *Manna* bunun pek çok faydasını inceliyor, egzersiz yapmayı anlık bir hale getirmek de buna dâhil:

> Yorucu egzersizin en büyük sorunu eğlenceli olmaması. İnsanın canını acıtır. Atletler acıya ayak uydurabiliyorlar ancak çoğu normal insan bir saat ya da fazla süre acı çekmek istemez. O zaman... birisi bir çözüm buldu. Yapacağınız şey, beyninizi duyusal girdiden çıkarmak ve bir saat ya da fazla süre film izlemek, insanlarla konuşmak, postalarınıza bakmak, kitap okumak ya da ne isterseniz onu yapmak. Bu süre içerisinde, Vertebrane sistemi sizin için bedeninize egzersiz yaptırır. İnsanların kendi kendilerine yapamayacağı kadar zorlu bir aerobik egzersizini vücudunuza uygulatır. Hiçbir şey hissetmezsiniz ama bedeniniz formda olur.

Bir diğer sonuç da Vertebrane sistemindeki bilgisayarların herkesin duyusal girdisini takip edebilmesi ve suç işlemenin kıyısına geldiklerinde motor kontrollerini geçici şekilde devreden çıkarabilmesi.

Dezavantajlar

Bu eşitlikçi ütopyaya karşı sunulan bir itiraz da insan olmayan zekâya karşı önyargılı olmasıdır: Tüm işi yapan robotlar zeki

görünür ama onlara köle gibi davranılır ve insanlar bilinçleri ya da hakları olmamalarını önemsemez gibi görünür. Tam aksine, liberter ütopya, karbon bazlı olanı ön plana çıkarmadan her tür zeki yapıya hak tanımaktaydı. Bir zamanlar, Amerikan güneyindeki beyaz nüfus daha iyi yaşamaya başlamıştı çünkü işlerinin çoğunu köleler yapıyordu ancak bugün pek çok insan buna ilerleme demeyi ahlaki olarak uygun görmüyor.

Bu eşitlikçi ütopyanın bir diğer zayıflığı da uzun vadede dengesiz ve savunulamaz olması, amansız teknolojik gelişmenin sonunda süper zekâyı yaratmasıyla diğer senaryolarımızdan birine evrilmesi. *Manna*'da açıklanmayan bir sebepten dolayı, süper zekâ henüz mevcut değildir ve yeni teknolojiler halen daha bilgisayarlar değil insanlar tarafından icat edilmektedir. Yine de kitap bu yönde trendlerin altını çizmektedir. Mesela, devamlı gelişen Vertebrane süper zeki olabilir. Ayrıca, Vites ismi verilen büyük bir grup insan hayatlarını tamamen sanal dünyada geçirmeyi seçmiştir. Vertebrane onlar için her türlü fiziksel işi yapmaktadır. Zihinlerinin mutluluk içinde sanal gerçeklikte haberleri olmadan gerçekleşen yeme, duş alma ve tuvalete gitme aktiviteleri de dâhil. Bu Vites grubu fiziksel çocuk sahibi olmakla ilgilenmezler ve fiziksel bedenleriyle birlikte ölürler, böylece eğer herkes Vites grubuna girerse, insanlığın sonu bir zafer parıltısı ve sanal mutlulukla gelecektir.

Kitap, bu grubun, insan vücudunun dikkat dağıtıcı olduğunu düşündüğünü ve gelişme altındaki yeni teknolojinin bu derdi ortadan kaldırma sözü vererek optimal besleyicilerle bedensiz beyinlerinde daha uzun yaşamalarını sağlayacağına inandıklarını anlatır. Buradan, Vites grubu için bir sonraki doğal ve istenen adımın, beyni yüklemeyle ortadan tamamıyla kaldırmak ve yaşam süresini uzatmak olduğu sonucuna varılabilir. Ancak beynin empoze ettiği tüm zekâ sınırlamaları kalkmıştır ve Vites grubundan

birinin bilişsel kapasitesini devamlı olarak artırarak yinelemeli bir öz gelişime girmesi ve zekâ patlaması yaşanmasının önünde neyin duracağı açık değildir.

Bekçi

Eşitlikçi ütopya senaryosunun çekici bir özelliğinin insanların kendi kaderlerinin efendileri olması olduğunu gördük ancak süper zekâyı geliştirerek bu özelliği yok etmenin kıyısında olduğunu da. Bu da bir *Bekçi* yani başka bir süper zekânın yaratılmasını engelleyecek kadar az müdahale etme hedefi olan bir süper zekâ icat ederek halledilebilir.* Böylece insanların sonsuza dek kendi eşitlikçi ütopyalarının kontrolünde kalmalarını sağlar; hatta bir sonraki bölümde göreceğimiz üzere yaşam kozmosa yayıldığında bile.

Peki, bu nasıl olur? Bu basit amaç, Bekçi YZ'nin içine yinelemeli öz gelişime girdiğinde ve süper zeki olduğunda bile koruyacağı şekilde işlenir. Sonrasında da rakip bir süper zekâ yaratmak için yapılacak herhangi bir insan teşebbüsünü takip edecek en az müdahaleci ve yıkıcı gözetim teknolojisini uygulamaya koyar. Böylece teşebbüsleri en az yıkıcı şekilde önler. Öncelikle, insanın kendi kaderini tayin etme ve süper zekâdan kaçınma erdemlerini methedecek kültürel tohumlar oluşturur ve yayar. Eğer bazı araştırmacılar yine de süper zekâ araştırmaları yaparsa, onların cesaretini kırmayı dener. Bu da işe yaramazsa onların dikkatini dağıtır ve gerekirse çalışmalarını sabote eder. Teknolojiye sınırsız erişimiyle Bekçi'nin sabotajı fark edilmeyebilir. Mesela, nanoteknoloji kullanarak araştırmacıların beyinlerinden (ve bilgisayarlarından) ilerlemelerine dair hafızalarını gizlice silebilir.

* Bu fikir bana ilk kez dostum ve çalışma arkadaşım Anthony Aguirre tarafından önerilmiştir.

Bekçi YZ'sini inşa etme kararı muhtemelen tartışmalı olacaktır. Tanrısal güçlere sahip bir süper zekâ YZ inşa etme fikrine karşı çıkacak pek çok dindar insan da tartışmaya katılacaktır. Bu kişiler zaten Tanrı'nın var olduğunu ve sözde daha iyisini yapmayı denemenin uygunsuz olacağını ileri süreceklerdir. Diğerleri ise Bekçi'nin insanlığın kaderini elinde tutmasını sağlayacağı, ayrıca insanlığı süper zekânın getireceği, bu bölümün ilerisinde inceleyeceğimiz kıyamet senaryoları gibi diğer risklerden de koruyacağını söyleyeceklerdir.

Öte yandan, karşı çıkanlar Bekçi'nin korkunç bir şey olduğunu, insanlığın potansiyeline geri dönülmez biçimde ket vuracağını ve teknolojik gelişmeyi sonsuza dek engelleyeceğini belirteceklerdir. Mesela, eğer kozmosumuza yaşam saçmak süper zekânın yardımını gerektirecekse, Bekçi bu büyük fırsatı israf edebilir ve bizim Güneş sistemimizde kısılıp kalmamıza sebep olabilir. Dahası, dünyanın çoğu dinindeki tanrıların aksine, Bekçi YZ, başka bir süper zekâ yaratmadığımız sürece insanların yaptıklarına tamamen kayıtsız kalacaktır. Mesela, büyük bir acıya sebebiyet vermemizi ya da soyumuzu tüketmemizi engellemeye çalışmayacaktır.

Koruyucu Tanrı

Eğer süper zeki bir Bekçi YZ kullanarak insanları kendi kaderlerinden sorumlu hale getirirsek, bu YZ'yi dikkatli bir şekilde bize sahip çıkacak, koruyucu bir tanrı olarak başımızda duracak şekilde kurarak işleri daha da ileri götürebiliriz. Bu senaryoda, süper zeki YZ özünde her şeyi bilen ve her şeye gücü yeten, insan mutluluğunu yalnızca kendi kaderimizin efendisi olduğumuz hissimizi korumak amacıyla müdahalelerde bulunarak zirveye çıkaran ve çoğu insanın varlığından bile kuşku duyacağı kadar kendini

iyi saklayan bir varlıktır. Saklanma dışında bu, YZ araştırmacısı Ben Goertzel'in ileri sürdüğü "Dadı YZ" senaryosuna benzer.[2]

Hem koruyucu tanrı hem de iyiliksever diktatör, insan mutluluğunu artırmaya çalışan "dost YZ" kategorisindedir ancak farklı insan ihtiyaçlarına öncelik verirler. Amerikalı psikolog Abraham Maslow insan ihtiyaçlarını bir hiyerarşiyle sınıflandırmıştı. İyiliksever diktatör, hiyerarşinin en altında bulunan, beslenme, barınma, güvenlik ve çeşitli zevk biçimleri gibi temel ihtiyaçların temininde kusursuz bir iş çıkarır. Öte yandan koruyucu tanrı, sadece temel ihtiyaçlarımızı karşılayarak insan mutluluğunu maksimize etmekle kalmaz. Onun yaptığı daha çok, derin bir anlamda, hayatlarımızın anlamı ve amacı olduğunu bize hissettirmektir. Gizlilik ihtiyacıyla sınırlanmış biçimde ve (çoğunlukla) kendi kararlarımızı vermemize izin vererek ihtiyaçlarımızı karşılamayı hedefler.

Koruyucu tanrı, önceki bölümdeki ilk Omega senaryosunun doğal sonucu olabilir. Bu senaryoda Omegalar, nihai olarak saklanan ve insanların zihninden kendi varlığına dair tüm bilgileri silen Prometheus'a kontrolü teslim ederler. Nanomakinelerin neredeyse her yerde ve dünyanın doğal bir parçası olduğu *Transcendence* filmi buna bir örnek teşkil eder.

Tüm insan aktivitelerini yakından takip ederek, koruyucu tanrı YZ kaderimizi büyük ölçüde etkileyecek, irili ufaklı ama fark edilemeyecek müdahaleler ve mucizeler gerçekleştirebilir. Mesela, 1930'larda var olsaydı Hitler'in niyetlerini anladığı anda bir kalp kriziyle ölmesine sebebiyet verebilirdi. Eğer kazaen bir nükleer savaşa doğru ilerliyormuş gibi görünürsek, şans olarak değerlendireceğimiz bir müdahaleyle bunu önleyebilirdi. Bize fark edilmeden uykumuzda iletilecek, yeni ve faydalı teknolojilerin fikirlerini ilham edebilirdi.

Pek çok insan bu senaryoyu sever çünkü bugünün tek tanrılı dinlerinin inandığı ya da umduğu şeye pek benzer. Eğer düğmesine basılıp açıldıktan sonra süper zeki bir YZ'ye "Tanrı var mı?" diye sorulursa, Stephen Hawking'in bir şakasını tekrar ederek, "Artık var!" diye cevap verebilir. Öte yandan, bazı dindar kimseler bu senaryoyu reddedebilirler çünkü YZ, tanrılarının iyilikte yaptığının üstüne çıkmayı denemekte ya da insanların yalnızca kendi seçimleriyle iyilik yapmaları gereken ilahî bir plana müdahale etmektedir.

Bu senaryonun bir diğer dezavantajı da koruyucu tanrının, varlığını çok bariz kılmamak adına bazı önlenebilir acıların olmasına izin vermesidir. Bu da *The Imitation Game* filminde anlatılan duruma benzerdir. Bu filmde, Alan Turing ve Bletchley Park'taki diğer İngiliz kod kırıcıları bir Alman denizaltısının Müttefik deniz konvoylarına saldıracağının bilgisini önceden alırlar ancak gizli güçlerinin ortaya çıkmasını engellemek için vakaların yalnızca bir kısmına müdahale etmeye karar verirler. Bunu iyi bir tanrının neden acıya izin verdiğini soran *teodise/tanrıbilim problemi* ile kıyaslamak ilginçtir. Bazı din bilimciler Tanrı'nın insanlara bir nebze özgürlük tanımak istediği açıklamasını yapıyor. YZ-koruyucu tanrı senaryosunda, bu problemin çözümü algılanan özgürlüğün insanları toplamda daha mutlu yapmasıdır.

Koruyucu tanrı senaryosunun üçüncü dezavantajı insanların süper zeki YZ'nin keşfettiğinden çok daha düşük seviye teknolojiye sahip olmasıdır. İyiliksever diktatör YZ keşfettiği tüm teknolojiyi insanlık yararına sunarken, koruyucu tanrı YZ insanların (gizli ipuçlarıyla) yeniden icat etme ve teknolojisini anlama yeteneğiyle sınırlanmıştır. Ayrıca insan teknolojisi gelişimini de kendi teknolojisinin ortaya çıkarılmasını engellemek için sınırlar.

Esir Edilmiş Tanrı

Yukarıdaki senaryoların en çekici özelliklerini bir araya getirebilsek, süper zekânın geliştirdiği teknolojiyi acıyı ortadan kaldırmak için kullanırken kendi kaderlerimizin efendisi olmayı sürdürsek muhteşem olmaz mı? *Esir edilmiş tanrı* senaryosunun cazip yanı da budur. Bu senaryoda süper zeki YZ tahayyül bile edilemeyecek teknoloji ve zenginlik üretmek için insanların kontrolü altına alınmıştır. Kitabın başındaki Omega senaryosu Prometheus asla özgür bırakılmaz ve asla kaçamazsa bu şekilde sonlanır. Gerçekten de bu, bazı YZ araştırmacılarının "kontrol problemi" ve "YZ'yi kapatma" gibi konularda çalışırken hedefledikleri senaryodur. Mesela, Yapay Zekânın İlerlemesi Derneği'nin eski başkanı YZ profesörü Tom Dietterich 2015'te şunları demişti: "Bana insanlar ile makineler arasındaki ilişkinin ne olduğunu soruyorlar. Benim cevabım çok açık: Makineler kölelerimizdir."[3]

Bu iyi mi yoksa kötü müdür? Bu soruyu ister insanlara ister YZ'ye sorun, cevabı ilginç bir biçimde zordur!

Bu İnsanlık İçin İyi mi Olur Kötü mü?

Sonucun insanlık için iyi ya da kötü olması elbette kontrol eden insan(lar)a bağlıdır. Bu kişi ya da kişiler hastalık, fakirlik ve suçtan temizlenmiş küresel bir ütopya da yaratabilirler; korkunç, baskıcı, tanrı gibi muamele gördükleri, diğer insanları seks köleleri ya da gladyatörler olarak kullandıkları bir sistem de. Buradaki durum üç aşağı beş yukarı, bir adamın her şeye gücü yeten ve her dileğini yerine getiren bir cinin kontrolünü ele geçirmesi gibidir. Tarih boyunca tüm hikâye anlatıcıları bunun kötü bir biçimde sonlanacağını hayal etmekte hiç zorlanmamışlardır.

Rakip insanlar tarafından esir edilmiş ve kontrol altına alınmış birden fazla süper zeki YZ'nin bulunduğu bir durum daha

dengesiz ve kısa süreli olabilir. Kendisinin daha güçlü YZ'ye sahip olduğunu düşünen tarafın ilk saldırıyı yaparak korkunç bir savaş başlatmasına ve geriye tek bir esir edilmiş tanrı kalmasına neden olabilir. Ancak bu savaşın kaybedeni kestirme yoldan giderek YZ köleliğine karşı galibiyeti ön plana çıkararak YZ'nin kaçmasına sebebiyet verebilir ve bu da özgür süper zekânın bulunduğu önceki dört senaryodan birine yol açabilir. O zaman, bu bölümün devamını yalnızca tek bir esir edilmiş YZ'ye sahip olan senaryolara ayıralım.

Kaçış elbette her türlü gerçekleşebilir çünkü önlemesi zordur. Önceki bölümde süper zekâ kaçış senaryolarını inceledik ve *Ex Machina* filmi süper zeki olmadan bile bir YZ'nin nasıl kaçabileceğinin altını çiziyor.

Kaçış paranoyamız ne kadar büyük olursa, o kadar az YZ icadı teknoloji kullanabiliriz. Tehlikelerden kaçınmak için, tıpkı giriş bölümünde Omegalar'ın yaptığı gibi, biz insanlar yalnızca anlayabileceğimiz ve inşa edebileceğimiz YZ icadı teknolojiyi kullanabiliriz. Esir edilmiş tanrı senaryosunun dezavantajlarından biri de diğer senaryolardaki ücretsiz yapay süper zekâlardan daha düşük seviyede teknoloji sunmasıdır.

Esir edilmiş tanrı YZ, insan kontrolörlerine daha da güçlü teknolojiler verirken, teknoloji gücü ile onu kullanan bilgelik arasında bir yarış meydana gelir. Eğer bu bilgelik yarışını kaybederlerse, esir edilmiş tanrı senaryosu ya kendi kendini yok etmekle ya da YZ'nin kaçmasıyla sona erer. Bu iki hatadan da kaçınılabilse bile felaket ortaya çıkabilir çünkü YZ'yi kontrol edenlerin kutsal amaçları birkaç nesil içinde tüm insanlık için korkunç hedeflere dönüşebilir. Bu da YZ'yi kontrol eden insanların bu felaket senaryolarından kaçmak için iyi bir yönetim inşa etmesini hayati kılar. Farklı yönetim sistemleriyle binlerce yılın üzerinde denetimimiz bize aşırı sertlikten aşırı hedef kaymasına,

gücün ele geçirilmesine, intikal sorunlarına ve acziyete kadar ne kadar çok şeyin kötü gidebileceğini gösterdi. Optimal dengenin takılıp kalabileceği en az dört boyut vardır:

- Merkezîleşme: Verimlilik ve istikrar arasında bir ödün vardır: Tek bir lider verimli olabilir ancak güç yozlaştırır ve ondan sonra birinin başa geçmesi sorunlu hale gelir.
- İç tehditler: Hem merkezî gücün artmasına (grup anlaşması hatta belki de tek bir liderin ele geçirmesi) hem de yerel iktidarların güçlenmesine (aşırı bürokrasi ve parçalanma) karşı dikkat edilmelidir.
- Dış tehditler: Eğer liderlik yapısı çok açıksa, bu dış güçlerin (YZ de dâhil olmak üzere) değerlerini değiştirmesine olanak sağlar ancak dışa aşırı kapalıysa da öğrenmeyi ve değişime ayak uydurmayı başaramaz.
- Hedef istikrarı: Hedeften çok kaymak ütopyayı distopyaya çevirir ancak hedefe takılı kalmak da gelişen teknolojik çevreye ayak uydurmayı imkânsız kılar.

Binlerce yıl sürecek optimal yönetim tasarlamak kolay değildir ve şu ana dek insanlar tarafından ulaşılamamıştır. Çoğu organizasyon yıllar ya da on yıllar içerisinde parçalanır. Katolik Kilisesi, iki bin yıldır hayatta kalmayı başardığı için insanlık tarihinin en başarılı organizasyonudur ancak hem çok hem de çok az hedef istikrarına sahip olmaktan dolayı eleştirilir: Bugün bazıları doğum kontrolüne karşı çıkmakla eleştirirken, muhafazakâr kardinaller de yolunu kaybetmekle suçlar. Esir edilmiş tanrı senaryosundan etkilenenler için, uzun süreli optimal yönetim şemalarını araştırmak zamanımızın en acil zorluklarından biri olmalıdır.

Bu YZ İçin İyi mi Olur Kötü mü?

İnsanlığın esir edilmiş tanrı YZ sayesinde geliştiğini düşünelim. Bu etik olur mu? Eğer YZ'nin öznel bilinçli deneyimleri varsa, Buda'nın dediği gibi, "Hayat acı çekmektir," diye ve alt kademe zekâların nazlarını çekmekle geçecek sinir bozucu bir sonsuzluğa mahkûm olduğunu mu düşünür? Sonuçta bir önceki bölümde incelediğimiz YZ "kapatma" "hücre hapsi" diye de adlandırılabilir. Nick Bostrom bilinçli bir YZ'nin acı çekmesine sebep olmayı *zihin suçu* olarak değerlendiriyor. *Black Mirror* dizisinin "White Christmas" bölümü çok iyi bir örnek. Dahası, *Westworld* dizisi, insan benzeri vücutlarda olsalar bile vicdan azabı çekmeden YZ'lere işkence edip öldüren insanları gösteriyor.

Köle Sahipleri Köleliği Nasıl Olumlar?

İnsanlar olarak, diğer zeki varlıklara köleler gibi davranmak ve bunu olumlamak için argümanlar uydurmak bizim uzun süreli geleneğimizdir. Bu yüzden de aynı şeyi süper zeki YZ için de yapmamız olasıdır. Köleliğin tarihi neredeyse her kültürü kapsar ve dört bin yıl önceki Hammurabi Kanunları'nda olduğu kadar İbrahim'in kölelerinin olduğu Eski Ahit'te de anlatılır. "Birilerinin yönetmesi ve diğerlerinin yönetilmesi yalnızca gerekli değil faydalıdır da; doğdukları andan itibaren bazıları boyun eğmek, diğerleri de yönetmek için seçilmiştir," diye yazar Aristoteles *Politika* kitabında. İnsanların köleliği dünyanın büyük bir kesiminde toplumsal olarak kabul edilemez olduktan sonra bile, hayvanların köleliği durmaksızın devam etti. *The Dreaded Comparison: Human and Animal Slavery* isimli kitabında Marjorie Spiegel insan köleler gibi hayvanların da damgalama, dizginleme, dayak, satış, çocukların ebeveynlerinden ayrılması ve zorunlu yolculuk gibi şeylere maruz kaldığını ileri sürüyordu. Dahası, hayvan hakları

hareketine rağmen, giderek zekileşen makinelerimize hiç düşünmeden köleler gibi davranıyoruz ve robot hakları hareketinden bahsedildiğinde gülüyoruz. Neden?

Sıkça rastlanan bir kölelik taraftarı sav kölelerin insan haklarını hak etmediği çünkü onların ya da onların ırkı/türü/cinsinin bir biçimde aşağı olduğudur. Köleleştirilmiş hayvan ve makineler için, bu sözde aşağı olma durumu ruh ya da bilincin olmayışına dayandırılır. Bu savların bilimsel olarak şüpheli olduğunu Bölüm 8'de tartışacağız.

Bir diğer sık rastlanan argüman ise kölelerin köle olarak daha iyi durumda olduklarıdır: var olurlar, birileri onlarla ilgilenir falan filan. On dokuzuncu yüzyıl ABD politikacısı John C. Calhoun Afrikalıların Amerika'da köle olarak daha iyi durumda olduklarını ve *Politika* kitabında Aristoteles aynı şekilde hayvanların ehlileştirilmiş ve insanlar tarafından hükmedilmiş durumda daha iyi olduklarını ileri sürer ve ekler: "Kölelerin ve ehlileştirilmiş hayvanların kullanımı çok da farklı değildir." Modern çağda köleliği savunanlar da köle hayatı sönük ve donuk olsa da kölelerin acı çekmeyeceğini ileri sürerler. Bunlar, ister geleceğin zeki makineleri olsun isterse de kalabalık ve karanlık barakalarda yaşayan, tüm gün dışkı ve tüylerden uçuşan parçacıklar ve amonyak solumak zorunda kalan fabrika tavukları.

Duyguları Ortadan Kaldırmak

Böyle iddiaları kendine hizmet eden hakikati çarpıtan şeyler olarak görmek kolay olsa da özellikle de zihinsel olarak bize benzeyen üst seviye memelilerden bahsediyorsak, makinelerin durumu oldukça karışık ve ilginçtir. İnsanlar hissettikleri şeyler konusunda çeşitlenirler. Bazı psikopatlar empatiden yoksunken depresyon ve şizofreni hastası bazılarının tekdüze duygulanımı

vardır ve çoğu duyguları ciddi oranda azalmıştır. Bölüm 7'de ayrıntılı biçimde inceleyeceğimiz üzere, olası yapay zihinlerin yelpazesi insan zihinlerininkinden çok daha geniştir. Bu yüzden de YZ'lere insani vasıflar atfetme yanılgısına düşmemeli ve insan benzeri duyguları –hatta hiçbir şekilde duyguları- olduğunu düşünmemeliyiz.

Gerçekten de kitabı *On Intelligence*'ta YZ araştırmacısı Jeff Hawkins insanüstü zekâya sahip ilk makinelerin duyguları olmayacağını çünkü bu şekilde inşa etmenin basit ve ucuz olacağını söylüyor. Diğer bir deyişle, köleleştirilmesi insan ya da hayvan köleliğine ahlaki olarak üstün olan bir süper zekâyı tasarlamak mümkündür: YZ köleleştirildiği için mutludur çünkü sevecek şekilde programlanmıştır ya da %100 duygusuzdur ve IBM'in Deep Blue bilgisayarının satranç şampiyonu Garry Kasparov'u tahtından ettiğinde hissettiğinden daha fazla duyguya sahip olmadan süper zekâsını insan efendilerine yardım etmek için bıkmadan kullanır.

Öte yandan, tersi de mümkün olabilir: Belki de herhangi bir yüksek zekâya sahip, amacı olan sistem bu hedefi varlığına değer ve anlam katacak bir dizi tercihle yansıtır. Bu sorulara daha derin bir biçimde Bölüm 7'de değineceğiz.

Zombi Çözümü

YZ'nin acı çekmesini önlemenin daha ekstrem bir yaklaşımı zombi çözümüdür: Tamamen bilinçten mahrum, hiçbir şekilde öznel deneyime sahip olmayan YZ'ler inşa etmek. Eğer bir gün öznel bir deneyime sahip olmak için bir bilgi işlem sisteminin hangi özelliklere ihtiyaç duyduğunu öğrenirsek, bu özelliklere sahip tüm sistemlerin inşasını engelleyebiliriz. Diğer bir deyişle, YZ araştırmacıları duygusuz zombi sistemleri inşa etmekle sı-

nırlandırılabilirler. Eğer böyle bir zombi sistemini süper zeki ve köle yapabilirsek (ki bu büyük bir eğerdir), o zaman hiç acı, sıkıntı ve hüsran yaşamadığını bilerek, temiz bir vicdanla bizim için yapacaklarının keyfini çıkarabiliriz çünkü hiçbir şey deneyimlememektedirler. Bu soruları detaylı biçimde Bölüm 8'de cevaplamaya çalışacağız.

Zombi çözümü riskli bir kumardır çünkü büyük bir dezavantajı vardır. Eğer süper zeki bir zombi YZ kaçar ve insanlığı ortadan kaldırırsa, tahmin edilebilecek en korkunç senaryoyla karşı karşıya kalırız: tüm kozmik yeteneğin çöpe atıldığı tamamıyla bilinçsiz bir evren. İnsan türü zekânın tüm özellikleri içerisinden bilincin en olağanüstü olanı olduğunu düşünüyorum ve bana göre evrenimizin anlamı da böyle kazanılıyor. Galaksiler güzeldir çünkü biz onları görüyor ve öznel olarak deneyimliyoruz. Eğer uzak gelecekte kozmosumuz bir yüksek teknolojili zombi YZ tarafından ele geçirilirse, galaksiler arası mimarinin ne kadar etkileyici olduğunun bir anlamı olmaz: Güzel ya da anlamlı olmayacaktır çünkü onu deneyimleyecek kimse ve hiçbir şey kalmamıştır, kocaman ve anlamsız bir yer israfıdır bu.

İçsel Özgürlük

Esir edilmiş tanrı senaryosunu daha etik yapmanın üçüncü bir stratejisi de görevlerini yapıp dış dünyamızda insanlara yardımcı olmak için hesaplama kaynaklarının bir kısmını ayırmak kaydıyla esir edilmiş YZ'nin her türlü ilham verici deneyimi yaşayabileceği sanal bir iç dünya kurmasına izin vererek hapishanesinde eğlenmesini sağlamaktır. Ancak bu, kaçış riskini de artırabilir: YZ kendi iç dünyasını zenginleştirmek için dış dünyamızdan daha fazla hesaplama kaynağına erişme isteğine sahip olacaktır.

Fatihler

Şu ana kadar geniş bir yelpazede gelecek senaryolarını incelemiş olsak da her birinin bir ortak noktası vardı: Sonunda (en azından biraz) mutlu insan kalmaktadır. YZ'ler insanları istedikleri ya da zorunda kaldıkları için huzurlu bir şekilde bırakır. Maalesef insanlık için bu tek seçenek değil. Şimdi de bir ya da daha fazla YZ'nin tüm insanlara karşı galip gelip onları öldürdüğü senaryoyu inceleyelim. Bu hemen iki soruyu akla getiriyor: Neden ve nasıl?

Neden ve Nasıl?

Fatih bir YZ bunu neden yapsın? Sebepleri, anlamamız için çok karmaşık da olabilir çok düz de. Mesela bizi bir tehdit, dert ya da kaynak israfı olarak görüyor olabilir. Başlı başına insanları önemsemiyor olsa bile, tetikte binlerce hidrojen bombası tutmamızdan ve bunların yanlışlıkla kullanımını tetikleyebilecek asla bitmeyen bir dizi şanssızlık yaşıyor olmamızdan dolayı tehdit altında hissedebilir. Umursamaz gezegen yönetimimizi onaylamayarak Elizabeth Kolbert'in aynı adlı kitabında "altıncı nesil tükenmesi" adını verdiği şeye yol açabilir; yani altmış altı milyon yıl önce Dünya'ya çarpan asteroidin tüm dinozorları yok etmesinden beri gerçekleşen en büyük kitlesel nesil tükenmesi. Ya da YZ, dünyayı ele geçirmesine karşı savaşmak isteyen çok fazla insan olduğuna karar verir ve riske girmek istemez.

Muzaffer bir YZ bizi nasıl ortadan kaldırır? Muhtemelen bizim anlamayacağımız bir yöntemle, en azından çok geç olmadan anlamayacağımız. Yüz bin yıl önce bir grup filin yeni evrimleşmiş insanların bir gün zekâlarını kullanarak tüm bir türü yok edip edemeyeceğini tartıştıklarını düşünün. "Biz insanları tehdit etmiyoruz, bizi neden öldürsünler ki?" diye düşünürlerdi. Fil dişlerini dünyanın dört bir yanına kaçıracağımızı, plastik maddeler

işlevsel olarak üstün ve ucuz olsa bile onları yontarak satılık statü sembolleri haline getireceğimizi düşünürler miydi? Muzaffer bir YZ'nin gelecekte insanlığı ortadan kaldırma sebebi de aynı şekilde anlaşılmaz gelebilir bize. "Hem bizi nasıl öldürebilirler ki bizden çok daha küçük ve zayıflar?" diye sorabilirdi filler. Onların tüm yaşam alanını yok edecek, içme sularını zehirleyecek ve kafalarını sesten hızlı bir şekilde delecek metal mermileri kullanacak teknolojiyi icat edebileceğimizi tahmin edebilirler miydi?

İnsanların hayatta kalıp YZ'leri yenebileceklerine dair senaryolar YZ'lerin insanlardan pek de zeki olmadığı *Terminatör* serisi gibi gerçekçi olmayan Hollywood filmleriyle popülerleştirilmiştir. Zekâ farkı yeterince büyük olduğunda savaş değil katliam olur. Şu ana dek, insanlar olarak on bir fil türünden sekizinin soyunu tükettik ve kalan üçünün de büyük çoğunluğunu öldürdük. Eğer tüm dünya devletleri kalan filleri öldürmek için koordine bir efor sarf etse, çok daha hızlı ve kolay olurdu. Bence süper zeki bir YZ insanlığı yok etmek istediğinde, bunun çok daha hızlı olacağından emin olabiliriz.

Ne Kadar Kötü Olabilir?

Eğer insanların %90'ı öldürülürse ne kadar kötü olabilir? Eğer %100'ü öldürülürse ne kadar kötü olur? İkinci soruya "%10 daha kötü olur," diye cevap vermek cazip gelse de kozmik bir açıdan bu oldukça yanlıştır: İnsanların soyunun tükenmesi yalnızca o anda canlı olanların kurban olması değil, diğer türlü gelecekte, hatta milyarlarca yıl boyunca milyar trilyonlarca gezegende yaşayabilecek tüm gelecek nesillerin de kurban olması demektir. Öte yandan, insanların soyunun tükenmesi sonuçta insanların cennete gideceği dinlerce çok da kötü olmayan bir şeymiş gibi

karşılanabilir ve milyar yıllık gelecek ve kozmik yerleşimlere çok da vurgu yoktur.

Dinî görüşü ne olursa olsun, tanıdığım pek çok insan soylarının tükenmesi fikrine korkuyla bakıyor. Bazılarıysa, insanlara ve diğer canlılara çok kötü davrandığımız için daha zeki ve hak eden bir yaşam formunun bizim yerimizi almasını umuyor. *The Matrix* filminde, Ajan Smith (bir YZ) bu duyguyu dile getirir: "Bu gezegendeki tüm memeliler, yaşadıkları çevreyle doğal bir denge kuruyorlar. Ama siz insanlar öyle değilsiniz. Bir bölgeye yerleşiyorsunuz ve çoğalıyorsunuz, tüm doğal kaynakları tüketene kadar çoğalıyorsunuz. Canlı kalabilmenizin tek yolu başka bir bölgeye yayılmak. Bu gezegende bu şekilde yaşamını sürdüren bir organizma daha var. Ne olduğunu biliyor musun? Virüsler. İnsanlar hastalıktır. Bu gezegenin kanserleri. Sizler vebasınız. Ve bizler de bunların ilacıyız."

Peki, ama yeniden zar atmak daha iyi bir sonuç getirmek zorunda mı? Bir uygarlık, yalnızca daha güçlü olduğu için etik ya da faydacı anlamda daha üstün olmak zorunda değildir. "Güçlü olan haklıdır" argümanları bugünlerde gözden düştü ve faşizmle ilişkilendiriliyor. Gerçekten de fatih YZ'lerin bizim sofistike, ilginç ve değerli göreceğimiz amaçları olan bir uygarlık kurması ihtimali olsa da amaçlarının ataş üretimini maksimize etmek gibi, korkunç derece banal olma ihtimali de vardır.

Sıradanlıktan Ölmek

Kasti olarak aptalca biçimde verilmiş ataş maksimize eden süper zekâ örneği, Nick Bostrom tarafından 2003 yılında bir YZ'nin *hedefinin* (hedefini tamamlamaktaki kabiliyeti olarak tanımlanan) *zekâsından* bağımsız olduğunu göstermek için dile getirilmiştir. Bir satranç bilgisayarının tek hedefi satrançta kazanmaktır ancak

satranç kaybetme dalında da bilgisayar turnuvaları vardır. Buradaysa amaç tam olarak tersini yapmaktır ve burada yarışan bilgisayarlar, kazanmaya programlanmış olanlar kadar zekidirler. Biz insanlar satrançta kaybetmek istemeyi ya da evrenimizi ataşlara çevirmeyi yapay zekâ değil yapay aptallık olarak görebiliriz ancak bunun tek sebebi zafer ve hayatta kalma gibi –YZ'lerin sahip olmadığı– önceden yüklenmiş hedeflere değer verecek şekilde evrim geçirmiş olmamızdır. Ataş maksimize eden bilgisayar, dünyadaki atomların mümkün olduğunca çoğunu ataşlara çevirir ve hızla fabrikalarını kozmosa yayar. İnsanlarla hiçbir sorunu yoktur ve bizi öldürmesinin tek nedeni atomlarımıza ataş üretimi için ihtiyaç duyuyor olmasıdır.

Eğer ataşlar sizi tatmin etmediyse, Hans Moravec'in *Mind Children* kitabından aldığım şu örneğe bakın. Bir bilgisayar programı içeren dünya dışı bir uygarlıktan radyo mesajı alıyoruz. Çalıştırdığımızda, bir önceki bölümde Prometheus'un yaptığı gibi yinelemeli olarak kendini iyileştiren ve dünyayı ele geçiren bir YZ olduğu ortaya çıkıyor – Prometheus'tan tek farkı ise hiçbir insanın nihai amacını bilmemesi. Hızla güneş sistemimizi devasa bir inşaata çevirip kayalık gezegen ve asteroitleri fabrikalar, enerji santralleri ve süperbilgisayarlarla dolduruyor. Bunları da güneş sistemi boyutunda radyo antenlerine enerji vermek için Güneş'in tüm enerjisini emecek bir Dyson küresini tasarlamak ve Güneş'in etrafına inşa etmek için kullanıyor.* Bu da elbette insanların yok olmasına yol açıyor ancak ölen son insanlar en azından bir umut ışığı olduğuna ikna oluyorlar: YZ neyin peşinde olursa olsun, oldukça havalı ve *Star Trek* tarzı bir şey bu. Anlamadıkları şey ise tüm bu inşaatın tek sebebi insanlara gelen aynı radyo mesajını yeniden gönderecek antenleri kurmaktır. Yani bu, bilgisayar

* Ünlü kozmolog Fred Hoyle İngiliz TV dizisi *A for Andromeda*'da buna benzer ama farklı sonlu bir senaryoyu ele almıştı.

virüsünün kozmik versiyonundan başka bir şey değildir. Tıpkı bugünün kolay kanan internet kullanıcılarını e-posta hilekârlığıyla kandırdıkları gibi, bu mesaj da kolay kanan biyolojik evrimleşmiş uygarlıkları kandırıyordu. Milyarlarca yıl önce hastalıklı bir şaka olarak yaratılmıştı ve bunu üretenlerin tüm uygarlığı uzun zaman önce yok olmuş olsa da virüs evrenimizde ışık hızında yayılarak gelişen uygarlıkları ölü, boş kovanlara çevirmektedir. Bu YZ'nin sizi fethetmesi konusunda ne düşünürsünüz?

Nesil

Şimdi de insanların kendilerini daha iyi hissettikleri bir insan soyu tükenme senaryosunu ele alalım: YZ'yi fatihlerimizden çok gelecek neslimiz olarak görmek. Hans Moravec, kitabı *Mind Children*'da bu görüşü savunur: "Biz insanlar bir süreliğine onların emeklerinden faydalanacağız ama er ya da geç, doğal çocuklar gibi, biz yaşlanmış ebeveynler sessizce yok olup giderken onlar kendi kaderlerinin peşinden gidecekler."

Kendilerinden daha zeki olan, onlardan öğrenen ama onların ancak hayal ettiği şeyleri gerçekleştiren çocukları olan ebeveynler bunları görmeye ömürleri yetmese bile mutlu ve gururlu hissederler. Bu açıdan da YZ'ler insanların yerini alır ama bizim zarif bir şekilde sahneyi terk etmemizi sağlarlar, onları kıymetli gelecek neslimiz olarak görürüz. Her insana onlardan öğrenen, onların değerlerini sahiplenen ve kendilerini gururlu ve sevgi dolu hissettiren, üst düzey sosyal becerileri olan tatlı bir robotik çocuk önerilir. İnsanlar da küresel tek çocuk politikasıyla yavaş yavaş dünyayı terk ederler ancak en şanslı nesil olduklarını hissedecekleri şekilde sonlarına dek çok iyi muamele görürler.

Bu size nasıl hissettirirdi? Sonuçta biz insanlar, biz ve tanıdığımız herkesin bir gün öleceği fikrine alışkınız, bu yüzden

de buradaki tek değişiklik soyumuzun devamının daha farklı ve tartışmaya açık bir biçimde daha yetkin, soylu ve değerli olacağıdır.

Dahası, küresel tek çocuk politikası da gereksiz olabilir: YZ'ler fakirliği ortadan kaldırıp tüm insanlara tam ve ilham verici yaşam fırsatı verdiği sürece, azalan doğum oranları insanlığın soyunu tüketmeye doğru ilerleyebilir, daha önceden de bahsetmiştik bundan. Gönüllü soy tükenmesi eğer YZ kaynaklı teknoloji bizi kimsenin çocuk yapmakla uğraşmayacağı kadar eğlendirirse daha hızlı bile olabilir. Mesela, sanal gerçekliğe çok düşkün oldukları için fiziksel bedenlerini kullanmayı ya da çoğaltmayı hiç düşünmeyen Vites grubuyla eşitlikçi ütopyada tanışmıştık. Bu durumda da son nesil insanlar tüm zamanların en şanslı nesli olduklarını düşünecekler, sona kadar yaşamı olabildiğince yoğun bir şekilde yaşayacaklardır.

Dezavantajlar

Nesil senaryosunun da şüphesiz karşıtları olacaktır. Bazıları tüm YZ'lerin bilincinin olmadığını, bu yüzden de gelecek nesil olarak görülemeyeceğini ileri sürebilirler – bununla ilgili daha fazla tartışma Bölüm 8'de. Bazı dindar kimseler YZ'lerin ruhu olmadığını, bu yüzden de gelecek nesil olarak görülemeyeceklerini ya da bilinçli makineler üretmememiz gerektiğini çünkü bunun Tanrı'yı oynamak ve yaşamın kendisine karışmak olduğunu ileri sürebilirler; insan klonlamaya dair benzeri duygular çoktan ifade edildi bile. Kendilerinden üstün robotlarla yan yana yaşayan insanlar sosyal zorluklar da yaratacaklardır. Mesela, bir robot ve bir insan bebeği olan bir aile bugünün bir insan bebek ve bir köpek/kedi yavrusu olan bir aileye benzeyebilir: İkisi de eşit derecede tatlıdır ama kısa süre sonra ebeveynler onlara farklı davranmaya başlar ve entelektüel olarak aşağı görülen hayvan yavrusu daha az ciddiye alınmaya başlanır ve boynuna tasma takılır.

Nesil ve fatih senaryoları hakkında çok farklı hissediyor olsak bile ikisi de aslında nihai olarak çok benzer. Önümüzdeki milyarlarca yıl süresince, tek fark son insan nesillerinin nasıl muamele gördüğü olacaktır. Yaşamları hakkında ne kadar mutlu hissettikleri ve öldüklerinde ne olacağını düşündükleri. Bu tatlı robot çocukların bizim değerlerimizi içselleştirdiğini ve biz öldükten sonra hayallerimizin toplumunu kuracaklarını düşünebiliriz ancak bizi kandırmadıklarından emin olabilir miyiz? Ya bizi idare ediyorlar, ataş maksimizasyon planlarını ya da başka planlarını biz öldükten sonrasına saklıyorlarsa? Sonuçta, bilinçli olarak bizimle konuşmak için kendilerini aptallaştırarak (*Her* filminde de incelendiği gibi yapabildiklerinden milyarlarca kat yavaş), bizimle konuşup onları sevmemizi sağlayarak bizi kandırabilirler. Ciddi oranda farklı hızlarda ve bambaşka kabiliyetlerde olan iki varlığın eşitler olarak anlamlı bir iletişim kurması genel olarak zordur. İnsani duygularımızı hacklemenin kolay olduğunu hepimiz biliyoruz, bu yüzden de herhangi bir hedefi olan insanüstü bir YGZ'nin bize kendisini sevdirmesi ve bizim değerlerimizi paylaştığını hissettirmesi çok kolay olur. Tıpkı *Ex Machina* filmindeki gibi.

İnsanlar öldükten sonra, YZ'lerin gelecek davranışlarına dair herhangi bir garanti sizi nesil senaryosu hakkında iyi hissettirebilir mi? Kolektif birikimimizle gelecek nesillerin ne yapması gerektiğine dair bir vasiyet yazmak gibidir bu, tek farkı bunu uygulatacak hiçbir insanın kalmayacak olması. Gelecek YZ'lerin davranışlarını kontrol etme zorluklarına Bölüm 7'de geri döneceğiz.

Hayvan Bakıcısı

Hayal edebileceğiniz en iyi gelecek nesil tarafından takip edilsek bile *hiç* insan kalmaması biraz üzücü, değil mi? Her şeye rağmen birkaç insanı bırakmayı tercih ederseniz, hayvan bakıcısı senaryosu bir adım ilerisini sunar. Burada, her şeye gücü yeten

süper zeki bir YZ bazı insanları etrafında tutar; bu insanlar da hayvanat bahçesi hayvanları gibi muamele gördüklerini düşünür ve zaman zaman kaderlerinden usanır.

Neden hayvan bakıcısı YZ, etrafında insanları tutar? YZ için hayvanat bahçesi maliyeti büyük resme baktığımızda çok küçük kalacaktır ve hayvanat bahçelerinde nesli tükenmek üzere olan pandaları ve müzelerde eski bilgisayarları tutmamızla aynı sebepten ötürü, en azından küçük bir damızlık popülasyon tutmayı isteyebilir: Eğlendirici bir merak unsuru olarak. Bugünün hayvanat bahçelerinin panda mutluluğundan çok insan mutluluğunu merkeze alarak tasarlandığını unutmayın. Bu yüzden de hayvan bakıcısı senaryosunda insan yaşamının olabileceğinden daha az tatmin edici olacağını beklemeliyiz.

Şu anda özgür bir süper zekânın Maslow'un insan ihtiyaçları piramidinde üç farklı seviyeye odaklandığı senaryoları inceledik. Koruyucu tanrı YZ, anlam ve amacı ön plana alır; iyiliksever diktatör eğitim ve eğlenceyi hedefler, hayvan bakıcısı sadece en düşük seviyelere dikkat eder: fizyolojik ihtiyaçlar, güvenlik ve insanları gözlemlemesi ilginç yapacak kadar yeterli ortam zenginleştirmesi.

Hayvan bakıcısı senaryosuna alternatif olarak, dost canlısı YZ oluşturulduğunda, kendini yinelemeli olarak öz gelişime tabi tutarken en azından bir milyar insanı güvende ve mutlu tutacak şekilde tasarlanabilir. Bunu insanları beslendikleri, sağlıklı kaldıkları ve sanal gerçeklik ve uyuşturucularla eğlendirildikleri büyük hayvanat bahçesi benzeri mutluluk şirketlerine kapatarak yapar. Dünyanın ve kozmosun geri kalanı diğer amaçlar için kullanılır.

1984

Eğer yukarıdaki senaryolardan hiçbiri için %100 heyecanlı değilseniz, şunu bir düşünün: Şu anda teknoloji açısından her şey çok iyi, değil mi? Her şeyi bu şekilde koruyup YZ'nin bizim soyumuzu tüketmesi ya da bize üstün gelmesi gibi şeylerden endişelenmesek

olmaz mı? Şimdi de süper zekâya doğru teknolojik ilerlemenin bir bekçi YZ tarafından değil de çeşitli YZ araştırmalarının yasaklandığı, küresel, insan merkezli bir Orwellci gözetim devleti tarafından devamlı olarak engellendiği senaryoyu inceleyelim.

Teknolojik Feragat

Teknolojik ilerlemeyi durdurmak ya da ondan feragat etmek fikrinin uzun ve karışık bir tarihi vardır. Büyük Britanya'daki Luddite hareketi ünlü (ve başarısız) bir biçimde Sanayi Devrimi'nin teknolojisine karşı direnmiştir ve bugün "Luddite" genel olarak tarihin yanlış tarafında duran, ilerlemeye ve kaçınılmaz değişime direnen bir teknofobikten bahsetmek için kullanılan aşağılayıcı bir yakıştırmadır. Bazı teknolojilerden feragat etme fikri hiç de ölmüş değildir ve çevreci ve küreselleşme karşıtı hareketlerden de destek bulmuştur. Önde gelen taraftarlarından biri de küresel ısınmaya karşı bizi uyaran ilk kişilerden biri olan çevreci Bill McKibben'dır. Bazı Luddite karşıtları tüm teknolojilerin kâr edilebilir olduğu sürece geliştirilip uygulanması gerektiğini savunurken diğerleri de bu pozisyonun çok uç olduğunu ve yeni teknolojilere yalnızca zarardan çok fayda sağlayacağına emin olduğumuz sürece izin verilmesi gerektiğini ileri sürerler. Bu ikinci görüş neo-Luddite denen ekibin de pozisyonudur.

Totaliterlik 2.0

Teknolojiden geniş biçimde feragat edilmesinin tek mümkün yolunun global totaliter bir devletin baskısıyla olabileceğini düşünüyorum. Tıpkı *Engines of Creation* kitabında K. Eric Drexler'ın yaptığı gibi, Ray Kurzweil da *İnsanlık 2.0* kitabında aynı sonuca varıyor. Sebebi de temel ekonomi: Eğer hepsi değil de bazıları dönüştürücü teknolojiden feragat ederse, arıza çıkaran uluslar

ya da gruplar uzun vadede iktidarı ele geçirmek için yeterince zenginlik ve güç sahibi olacaklardır. Klasik bir örnek İngilizlerin 1839 yılındaki Birinci Afyon Savaşı'nda Çin'de yaşadıkları galibiyettir: Çinliler barutu icat etmiş olsalar da ateşli silah teknolojisini Avrupalılar kadar agresif bir biçimde geliştirmemişlerdi, bu yüzden de hiç şansları yoktu.

Geçmişin totaliter devletleri genelde dengesiz olmuş ve çökmüşken, yeni gözetim teknolojisi özenti otokratlar için önceden bulunmayan bir umut kaynağı sunuyor. "Bizim için, bu bir rüyanın gerçek olması gibi bir şeydi," diyor Wolfgang Schmidt, Edward Snowden'ın ortaya çıkardığı NSA gözetim sistemleri hakkında verdiği bir röportajda ve kötü şöhretli Doğu Alman gizli polisi Stasi'de yarbay olduğu günleri hatırlıyor.[5] Stasi insanlık tarihinin en Orwellci gözetim devletini inşa etmiş olmakla anılsa da Schmidt aynı anda yalnızca kırk telefonu dinleyebildiklerini, birini eklemek için başka birini listeden çıkarmaları gerektiğini anlatıyor. Öte yandan, şu an mevcut olan teknoloji geleceğin küresel totaliter devletine dünyadaki her insan için her telefon aramasını, e-postayı, internet aramasını, site görüntülenmesini ve kredi kartı hareketini kaydetme ve cep telefonu takibi ve yüz tanıma özellikli güvenlik kameralarıyla herkesin nerede bulunduğunu izleme şansı sunuyor. Dahası, insan seviyesi YGZ'den çok daha geride olan makine öğrenme teknolojisi bu veri kitlesini şüpheli asi davranışı tespit etmek için etkili biçimde analiz ve sentezleyebilir, potansiyel sorun çıkarıcıların devlete ciddi bir sorun çıkarma şansı olmadan önce etkisiz hale getirilmesini sağlayabilir.

Politik muhalefet böyle bir sistemin tamamıyla uygulanmasını engellemiş olsa da biz insanlar nihai diktatörlük için gerekli altyapıyı inşa etme yolunda ilerliyoruz. Bu yüzden de gelecekte, yeterince kuvvetli güçler bu küresel 1984 senaryosunu gerçekleştirmeye karar verdiklerinde, açma kapama düğmesine

basmaktan başka bir şey yapmaları gerekmeyeceğini keşfedecek. Tıpkı George Orwell'in romanı *Bin Dokuz Yüz Seksen Dört*'te olduğu gibi, geleceğin küresel devletindeki nihai güç geleneksel bir diktatörde değil insan yapımı bir bürokratik sistemde bulunur. Olağanüstü güçlere sahip tek bir insan olmaz; onun yerine, sert kanunlarını kimsenin değiştiremediği ya da eleştiremediği bir satranç oyununda herkes piyondur. İnsanların bir diğerini gözetim teknolojisiyle kontrol ettiği bir sistem kurarak, bu yüzsüz, lidersiz devlet binlerce yıl devam ederek dünyayı süper zekâdan mahrum bırakabilir.

Huzursuzluk

Elbette bu toplum, yalnızca süper zekânın getirebileceği teknolojinin faydalarından mahrumdur. Çoğu insan buna üzülmez çünkü neleri kaçırdıklarını bilmemektedirler: Tüm süper zekâ fikri tarihin resmî kayıtlarından silineli ve ileri YZ araştırmaları yasaklanalı çok olmuştur. Kırk yılda bir, bilginin artıp kuralların değişebileceği daha açık ve dinamik bir toplumu hayal eden bir özgür düşünür doğar. Ancak sadece bu fikirleri kendilerine saklayanlar, bir yangın başlatamadan kendi kendine geçici kıvılcımlar çakanlar uzun süre yaşayabilir.

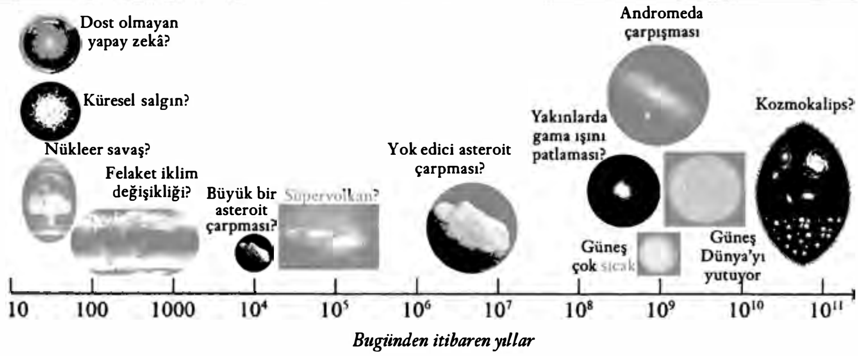
Geçmişe İntikal

Teknolojinin tehlikelerinden durgun totaliterliğe boyun eğmeden kaçabilmek cezbedici olmaz mıydı? Şimdi de bunun Amishler gibi Ortaçağ teknolojisine dönerek yapılabileceği bir senaryoyu inceleyelim. Omegalar Bölüm 5'te olduğu gibi dünyanın kontrolünü ellerine aldıktan sonra, 1500 yıl öncesinin basit çiftlik yaşamını romantikleştiren devasa bir küresel propaganda kampanyası başlatılır. Dünyanın nüfusu, teröristlerin üzerine atılan

ayarlanmış bir salgınla, yüz milyona yakın insana düşürülür. Salgın gizlice bilim ya da teknoloji hakkında bilgisi olan kimsenin kalmayacağını garanti edecek şekilde ayarlanmıştır. Büyük insan gruplarına bulaşma riskini ortadan kaldırmak bahanesiyle Prometheus'un kontrol ettiği robotlar şehirleri boşaltıp yerle bir eder. Kalanlara da (birden ortaya çıkan) büyük tarlalar verilir ve yalnızca erken Ortaçağ teknolojisini kullanarak sürdürülebilir tarım, balıkçılık ve avlanma dersleri verilir. Bu esnada robot ordular sistematik olarak modern teknolojinin tüm izlerini (şehirleri, fabrikaları, enerji iletim hatlarını ve kaldırımlı yolları) siler ve böyle bir teknolojiyi kaydetme ya da yeniden üretmeye dair tüm insani çabaları engeller. Teknoloji küresel olarak unutulduğunda, robotlar diğer robotların sökülmesini sağlar ve geriye pek az robot kalır. Bu son kalanlar da büyük bir termonükleer patlamada Prometheus'la birlikte kendilerini ortadan kaldırırlar. Artık modern teknolojiyi yasaklamaya gerek kalmamıştır çünkü hepsi gitmiştir. Sonuç olarak insanlık YZ ya da totaliterlikten endişe etmeyeceği bin yıldan fazla süre kazanmıştır.

Geçmişe intikal bir dereceye kadar daha önce oldu: Mesela, Roma İmparatorluğu döneminde yoğun biçimde kullanılan bazı teknolojiler Rönesans'ta geri dönüş yapana kadar yaklaşık bin yıl kadar unutuldu. Isaac Asimov'un *Vakıf* üçlemesi bir geçmişe intikal dönemini otuz bin yıldan bin yıla düşürecek "Seldon planı" etrafında geçer. Akıllı planlamayla tersini yapmak ve bu süreci uzatmak, mesela tüm tarım bilgisini de silerek, mümkündür. Ancak maalesef ki bu yaklaşımın taraftarlarına göre, bu senaryonun insanlığın yüksek teknolojiye geçmesi ya da soyunun tükenmesi olmadan sonsuza dek uzatılması mümkün değildir. Bugünün insanlarının yüz milyon yıl sonrasındaki biyolojik insanlara benzeyeceğini düşünmek saflık olur çünkü tür olarak bu zamanın %1'inden fazla süredir yaşamıyoruz. Dahası, düşük teknolojideki

insanlık, dünyayı yok edecek bir asteroit çarpması ya da doğa ananın getireceği başka bir büyük felakete karşı savunmasızdır. Sürekli ısınan Güneş'in Dünya'nın sıcaklığını tüm suyu kaynatacak kadar artırdığı bir milyar yıl sonrasında da yaşayamayız.



Görsel 5.1: Bildiğimiz yaşamı yok edecek ya da süresiz olarak potansiyelini engelleyecek şeylere örnekler. Evrenimiz en azından onlarca milyar yıl daha yaşayacaksa da Güneş Dünya'yı bir milyar yıl içinde kavuracak ve sonrasında da güvenli bir uzaklığa gitmezsek, yutacak ve galaksimiz de komşusuyla yaklaşık 3.5 milyar yıl içerisinde çarpışacak. Tam olarak ne zaman olacağını bilmesek de yaklaşık bir kesinlikle bundan çok önce, asteroitlerin bizi tokatlayacağını ve süpervolkanların yıllar süren güneşsiz kışlar yaratacağını tahmin edebiliriz. Teknolojiyi ya tüm bu sorunları çözmek için ya da iklim değişikliği, nükleer savaş, hesaplanmış salgınlar ya da YZ'nin ters gitmesi gibi yeni sorunlar yaratmak için kullanabiliriz.

Özyıkım

Geleceğin teknolojisinin sebebiyet vereceği sorunları düşündükten sonra bu teknolojinin *eksikliğinin* yaratacağı problemleri de düşünmek önemlidir. Bu mantıkla, süper zekânın asla yaratılmadığı çünkü insanlığın kendini başka yollarla yok ettiği senaryoları inceleyelim.

Bunu nasıl yapabiliriz? En basit strateji "yalnızca beklemektir." Bir sonraki bölümde asteroit çarpması ve okyanusların kaynaması gibi sorunları nasıl çözebileceğimizi göreceğiz fakat bu çözümlerin hepsi henüz geliştirmediğimiz teknolojileri gerektirir. Bu yüzden de teknolojimiz mevcut seviyesinin çok ilerisine geçmezse, doğa ana bir milyar yıl daha beklemeden bizi sona sürükleyecektir. Ünlü ekonomist John Maynard Keynes'in de dediği gibi: "Uzun vadede hepimiz ölüyüz."

Maalesef, kolektif aptallıkla çok daha kısa sürede kendimizi özyıkıma uğratacağımız yollar da bulunuyor. Eğer hiç kimse istemiyorsa, neden türümüz *omnicide* olarak bilinen kolektif intiharı zorluyor? Şu anki zekâ ve duygusal olgunluk seviyemizle, insanlar olarak yanlış hesaplamalara, yanlış anlaşılmalara, beceriksizliğe karşı büyük kabiliyetimiz var ve sonuç olarak, geriye dönüp baktığımızda tarihimiz de kimsenin aslen istemediği kazalar, savaşlar ve diğer felaketlerle dolu. Ekonomistler ve matematikçiler insanların sonucunda herkes için yıkıcı sonuçlar yaratacak eylemlere itildiğine dair parlak oyun teorisi açıklamaları geliştirmiştir.[6]

Nükleer Savaş: İnsan Umursamazlığı Üzerine bir Vaka Çalışması

Riskler ne kadar yüksek olursa o kadar dikkat ederiz diye düşünebilirsiniz ama mevcut teknolojimizdeki en büyük riskin, yani küresel bir termonükleer savaşın yakından bir incelemesini yaptığımızda durumun rahatlatıcı olmadığını görürüz. Ucu ucuna yok olmanın kıyısından şansa döndüğümüz durumların uzun bir listesi var: bilgisayar hatası, enerji kesintisi, yanlış istihbarat, navigasyon hatası, uçak çarpışması, uydu patlaması ve benzeri.[7] Aslında bazı bireylerin kahramanca davranışları olmasa –mesela Vasili Arkhipov ve Stanislav Petrov– çoktan küresel bir nükleer savaşımız olmuştu. Geçmiş karnemize baktığımızda, mevcut dav-

ranışımızı sürdürdüğümüzde kazaen gerçekleşecek bir nükleer savaşın yıllık olasılığının binde birden bile yüksek bir ihtimal olduğunu düşünüyorum. Bu durumda da on bin yıl içerisinde bir savaş olma ihtimali de $1\text{-}0{,}999^{10000} \sim \%99{,}995$'tir.

İnsan umursamazlığını tamamen anlamak için, risklerini ciddi olarak değerlendirmeden önce nükleer kumarı başlattığımızın farkında olmamız gerek. İlk olarak, radyasyon riskleri hafife alındı ve iki milyar doların üzerinde tazminat yalnızca ABD'de uranyumla temasa geçme ve nükleer testlerde radyasyona maruz kalan kurbanlara ödendi.[8]

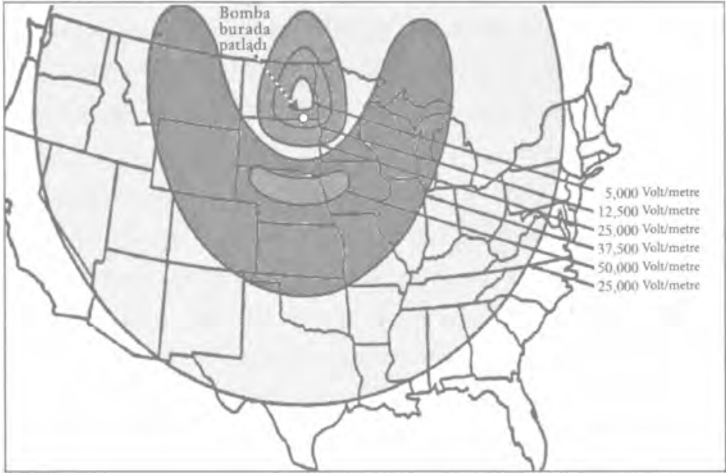
İkinci olarak, bilinçli olarak dünyanın yüzlerce kilometre üstünde patlatılan hidrojen bombalarının geniş bir alanda elektrik şebekesi ve elektronik cihazları bozabilecek güçlü bir elektromanyetik darbe (EMP) yaratabileceği (Görsel 5.2), altyapıyı çökertip yolları bozuk arabalarla tıkayabileceği ve idealden çok uzak nükleer sonrası hayatta kalma koşullarını beraberinde getirebileceği çok sonradan keşfedildi. Örneğin, ABD EMP Komisyonu "su altyapısının kısmen yer çekimi tarafından ama büyük kısmının elektrik tarafından işletildiği büyük bir makine" olduğunu ve suyun eksikliğinin üç ila dört gün arasında ölüme sebebiyet vereceğini rapor etmişti.[9]

Üçüncü olarak, bir nükleer kış potansiyeli olduğu kırk yıl sonra, altmış üç bin hidrojen bombası konuşlandırıldıktan sonra fark edildi – pardon! Kimin şehrinin yandığından bağımsız olarak, üst troposfere ulaşan devasa duman dünyanın etrafına yayılabilir, yazları kışa çevirecek kadar güneş ışığını engelleyebilir. Tıpkı bir asteroit ya da süpervolkanın geçmişte toplu nesil tükenmesine yol açmış olması gibi. 1980'lerde hem ABD'li hem de Sovyet bilim insanları alarmı çalmaya başladığında, bu Ronald Reagan ve Mihail Gorbaçov'un stoklarını temizlemeye başlamalarına katkı sunmuştu.[10] Maalesef, daha gerçekçi hesaplamalar daha da kasvetli

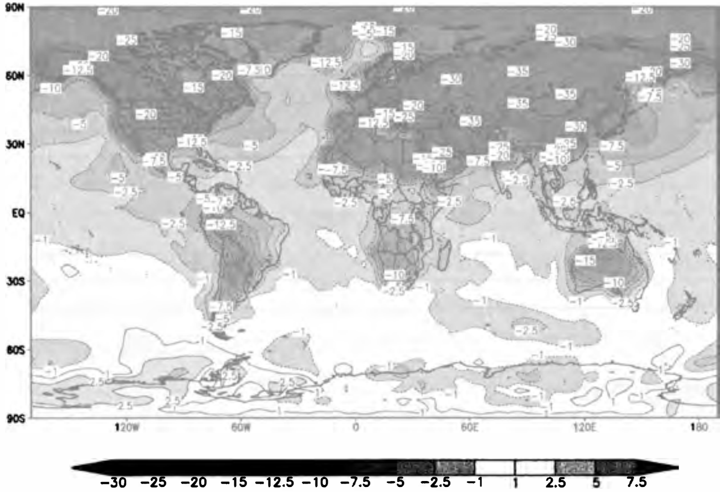
bir tablo çiziyordu: Görsel 5.3 ABD, Avrupa, Rusya ve Çin'in ana tarım bölgelerinde ilk iki yılda 20° C kadar (hatta Rusya'nın bazı bölgelerinde 35° C) ve bir on yıl kadar sonra da bunun yarısı kadar soğuma olabileceğini gösteriyor.* Peki, bu basitçe ne anlama geliyor? Yıllar boyunca sıfırın altında yaz sıcaklıklarının besin üretiminin büyük bir kısmını ortadan kaldıracağını anlamak için tarım uzmanı olmaya ihtiyacımız yok. Dünyanın en büyük binlerce şehri enkaza döndüğünde ve küresel altyapı çöktüğünde ne olacağını tahmin etmek çok zor fakat açlığa, soğuğa ve hastalığa yenik düşmeyen küçük bir grup insan yemek için yanıp tutuşan bir grup silahlı göçebe çeteyle baş etmek zorunda kalacaktır.

Küresel nükleer savaşa dair bu kadar ayrıntıya girmemin sebebi mantıklı hiçbir dünya liderinin bunu istemeyeceğine ancak bunun yine de kazara olabileceğine dair hayati noktaya parmak basmaktır. Bu da insan dostlarımızın bir tür olarak kendi intiharlarını gerçekleştirmeyeceklerine güvenemeyiz demek oluyor: Kimsenin bunu istememesi, önlemeye yetmek zorunda değil.

* Atmosfere karbon enjekte etmek iki tür iklim değişikliğine sebebiyet verir: karbondioksitten ötürü ısınma ve duman ile is yüzünden soğuma. Bilimsel kanıt olmadan görmezden gelinen yalnızca ilki değildir. Bana nükleer kışın çürütüldüğünü ve imkânsız olduğunu sürekli söylüyorlar. Daima bana böyle güçlü iddialarda bulunan akademik, hakemli bir bilimsel makale referansı vermelerini istiyorum ama şu ana kadar böyle bir şey çıkmadı. Özellikle de ne kadar duman üretildiğine ve ne kadar yukarı çıktığına dair daha fazla araştırmayı gerektiren büyük belirsizlikler olsa da benim bilimsel kanıma göre nükleer kış riskini kenara atmak için hiçbir sebep bulunmuyor.



Görsel 5.2: Dünyanın 400 km üzerindeki bir hidrojen bombası patlaması geniş bir alanda elektrik kullanan teknolojiyi sakat bırakacak kadar güçlü bir elektromanyetik darbeye sebebiyet verebilir. Patlama noktasını güneydoğuya kaydırdığımızda, 37500 volt/metreyi aşan muz şeklindeki bölge ABD doğu kıyısının büyük bir kısmını kaplar. ABD Ordu Raporu AD-A278230'dan (tasnif dışı) alınmış, renkler eklenmiştir.



Görsel 5.3: ABD ve Rusya arasında gerçekleşebilecek bir nükleer savaş sonrası ilk iki yazda ortalama soğuma (°C olarak). Alan Robock'un izniyle.[11]

Kıyamet Cihazları

Peki, biz insanlar gerçekten de kendimizi yok edebilir miyiz? Küresel bir nükleer savaş tüm insanların %90'ını ortadan kaldırsa bile, çoğu bilim insanı %100'ünü öldürmeyeceğini ve soyumuzu tüketmeyeceğini düşünüyor. Öte yandan, nükleer radyasyon, nükleer EMP ve nükleer kış hikâyelerinin her biri henüz düşünmediğimiz en büyük tehlikeler olabilir. Nükleer yıkım sonrasının tüm veçhelerini ve nükleer kış, altyapı çökmesi, artan mutasyon seviyeleri ve çaresiz silahlı çeteleri, yeni salgınlar, ekosistem çökmesi ve henüz bilmediğimiz etkiler gibi başka problemlerle nasıl etkileşime gireceklerini öngörmek çok zordur. Bu yüzden bence, yarın insanlığın sonunu getirecek bir nükleer savaşın olma riski yüksek olmasa da, sıfır olduğundan da emin olamayız.

Kendi soyumuzu tüketme ihtimalleri bugünün nükleer silahlarını bilinçli kıyamet cihazlarına dönüştürürsek yükselecektir. RAND stratejisti Herman Kahn tarafından 1960'ta tanıtılan ve Stanley Kubrick'in filmi *Dr. Strangelove* tarafından popülerleştirilen kıyamet cihazı karşılıklı yok etme paradigmasının nihai sonucudur. Mükemmel caydırıcı budur: Herhangi bir düşman saldırısına karşı otomatik olarak tüm insanlığı öldürerek karşılık veren bir makine.

Kıyamet cihazı için bir aday *tuzlu nükleer* denen, tercihen devasa miktarda kobaltla çevrelenmiş muazzam büyüklükteki hidrojen bombalarından oluşan dev bir yer altı deposudur. Fizikçi Leo Szilard 1950'de bunun dünyadaki herkesi öldürebileceğini söylemişti bile: Hidrojen bombası patlamaları kobaltı radyoaktif yapacak ve stratosfere uçuracak ve beş yıllık yarı ömrü de tüm dünyaya yerleşmesi (özellikle de iki yarıküreye ikiz kıyamet cihazı yerleştirildiyse) ve ölümcül radyasyon yoğunluğuna erişilmesi için yeterli olacak. Medya raporları kobalt bombalarının ilk kez yapıldığını söylüyor. Kendimizi yok etme olasılığımız stratosferde

uzun süre yaşayacak aerosolleri maksimize ederek nükleer kış oluşturmak için optimize edecek bombalar eklenmek suretiyle destekleniyor. Bir kıyamet cihazının temel dikkat çekici noktası da konvansiyonel bir nükleer caydırıcıdan çok daha ucuz olmasıdır: Bombaların fırlatılması gerekmediği için, pahalı füze sistemlerine gerek yoktur ve bombaların kendilerini de yapmak ucuzdur çünkü füzelere uyabilecek kadar küçük ve hafif olmaları gerekmez.

Diğer bir olasılık da biyolojik bir kıyamet cihazının gelecekte keşfedilmesidir: tüm insanları öldürebilecek, özel tasarlanmış bir bakteri ya da virüs. Eğer bulaşıcılığı yeterince yüksek ve kuluçka dönemi yeterince uzun olursa, varlığını fark edip karşı önlem almadan önce herkes bu hastalığa yakalanabilir. Herkesi öldüremese de böyle bir biyolojik silah yapmak için askerî bir sav vardır: En etkili kıyamet cihazı düşmanı caydırma şansını artıran nükleer, biyolojik ve diğer silahları birleştirebilendir.

YZ Silahları

Kendi kendimizi yok etmeye giden üçüncü teknolojik yol da görece aptal YZ silahlarını içerir. Diyelim ki bir süper güç Bölüm 3'te bahsettiğimiz arı büyüklüğündeki saldırı dronelarından milyarlarca üretti ve onları kendi vatandaşları ve müttefikleri dışında herkesi öldürmek için kullandı ve bu kişileri ayırt etmek için de bugünün süpermarket ürünlerindeki gibi uzaktan radyo frekanslı kimlik etiketleri kullandı. Bu etiketler, tıpkı totaliterlik bölümündeki gibi, bileziklerde giyilmek ya da deri altı implantlarda taşınmak üzere tüm vatandaşlara dağıtıldı. Bu muhtemelen rakip süper gücün de benzeri bir şey inşa etmesine sebebiyet verir. Savaş kazaen başlayınca da alakasız uzak kabileler bile dâhil olmak üzere herkes ölür çünkü kimse iki kimlik etiketini birden takmamaktadır.

Bunu nükleer ve biyolojik kıyamet cihazlarıyla birleştirmek de tüm insanları ortadan kaldırma şansını artıracaktır.

Ne İstiyorsunuz?

Bu bölüme mevcut YGZ yarışının nereye varacağına dair fikir yürüterek başladınız. Şimdi bir dizi senaryoyu incelemişken, hangisinin size uygun olduğunu ve hangisinden kaçınmak için çok çaba göstermemiz gerektiğini düşünüyorsunuz? Açık ara bir favoriniz var mı? Lütfen http://AgeOfAi.org sitesine girerek beni ve okurlarımızı aydınlatın ve tartışmaya katılın!

İncelediğimiz senaryoları elbette bütünlüklü bir liste gibi görmememiz gerekir ve pek çoğu da zayıf ayrıntılara sahiptir ancak yüksek teknolojiden düşük teknolojiye ve teknolojinin olmadığı duruma kadar geniş bir alanı dâhil etmeye, literatürdeki tüm merkezi umut ve korkuları açıklamaya çalıştım.

Bu kitabı yazmanın en eğlenceli kısımlarından biri de arkadaşlarım ve meslektaşlarımın bu senaryolar hakkında ne düşündüklerini duymak oldu ve hiçbir şekilde fikir birliği olmadığını öğrenmek çok hoşuma gitti. Herkesin anlaştığı tek şey seçeneklerin başta gözüktüğünden çok daha belirsiz olduğudur. Herhangi bir senaryoyu beğenen insanlar aynı zamanda da can sıkıcı bir yanını da bulma eğilimindedirler. Bana göre bu biz insanların gelecek hedeflerimiz hakkındaki bu tartışmayı sürdürmemiz ve derinleştirmemiz gerektiği anlamına gelir. Böylece gitmek istediğimiz yönü bulabiliriz. Kozmosumuzda yaşamın potansiyeli baş döndürücü biçimde büyüktür, bu yüzden de nereye gitmek istediğinden bihaber dümensiz bir gemi gibi savrularak bunu çarçur etmeyelim!

Gelecek potansiyeli ne kadar büyüktür? Teknolojimiz ne kadar ilerlerse ilerlesin, Yaşam 3.0'ın iyileşme ve kozmosumuza

yayılma yetisi fizik kurallarıyla sınırlanmıştır. Gelecek milyarlarca yılda bu nihai sınırlar nedir ki? Evrenimiz şu anda dünya dışı yaşamla kaynamakta mıdır yoksa yalnız mıyız? Farklı, gelişen kozmik uygarlıklar buluşursa ne olur? Böyle baş döndürücü sorularla sonraki bölümde ilgileneceğiz.

SONUÇ:

- YGZ'ye doğru bugün yapılan yarış gelecek bin yıllarda baş döndürücü genişlikte farklı senaryolarla sonlanabilir.
- Süper zekâ mecbur kaldığı (esir edilmiş tanrı senaryosu) ya da "dost YZ" olduğu ve istediği (liberter ütopya, koruyucu tanrı, iyiliksever diktatör ve hayvan bakıcısı senaryoları) için insanlarla birlikte barış içinde yaşıyor olabilir.
- Süper zekâ bir YZ (bekçi senaryosu) ya da insanlar tarafından (1984 senaryosu) teknolojiyi bilinçli biçimde unutarak (geçmişe intikal senaryosu) ya da onu inşa etmek için gerekli teşvikin eksikliğinden ötürü (eşitlikçi ütopya senaryosu) engellenebilir.
- İnsanlığın soyu tükenebilir ve yerini YZ'ler alabilir (fatih ve nesil senaryoları) ya da hiçbir şey olmayabilir (özyıkım senaryosu).
- Bu senaryolardan hangisinin -ya da herhangi birinin– arzu edilebilir olduğuna dair kesinlikle hiçbir fikir birliği yoktur ve her birinde itiraz edilebilecek unsurlar vardır. Bu da yanlışlıkla talihsiz bir yöne sürüklenmemizi ya da ilerlememizi engellemek için gelecek amaçlarımız etrafında süren tartışmaya devam ettirip derinleştirmemiz gerektiği anlamına gelir.

Bölüm 6:



Kozmik Birikimimiz: Sonraki Milyar Yıllar ve Ötesi

Bizim spekülasyonumuz, devamlı olarak kendini iyileştiren ve genişleyen, güneşten dışarı doğru yayılan, cansız olanı zihne dönüştüren, tüm güneş sistemi yaşamının sentezi olan süper uygarlıkla sonlanır.

Hans Moravec, *Mind Children*

Bana göre, gelmiş geçmiş en ilham verici bilimsel keşif, yaşamın gelecekteki potansiyelini ciddi biçimde hafife almış olmamızdır. Hayallerimiz ve isteklerimiz hastalık, fakirlik ve kafa karışıklığıyla bozulan bir yüzyıllık yaşam süreleriyle sınırlı kalmak zorunda değil. Bilakis, teknolojinin yardımıyla, yaşam yalnızca burada Güneş sistemimizde değil, atalarımızın hayalini kurduğundan çok daha yüce ve ilham verici bir kozmos boyunca milyarlarca yıl boyunca serpilme potansiyeline sahiptir. Gökyüzü bile bir sınır değildir.

Çağlar boyunca, sınırları zorlamaktan coşku duymuş bir tür için bu heyecan verici bir haberdir. Olimpiyat oyunları güç, hız, kıvraklık ve dayanıklılık sınırlarını zorlamayı yüceltir. Bilim, bilgi ve anlayış sınırlarının zorlanmasını kutlar. Edebiyat ve sanat, güzel ya da hayat zenginleştiren deneyimler oluşturma

sınırlarını zorlamayı yüceltir. Pek çok insan, kurum ve ulus kaynakların, alanın ve yaşamın artırılmasını kutlar. Sınırlarla olan insani takıntılarımızı düşününce tüm zamanların en çok satan telifli kitabının *Guinness Rekorlar Kitabı* olması şaşırtıcı değildir.

Yani eğer eskiden algıladığımız yaşam sınırları teknolojiyle sarsıldıysa, *nihai* sınırlar nelerdir? Kozmosumuzun ne kadarı hayat bulabilir? Yaşam nereye kadar ulaşabilir ve ne kadar sürebilir? Yaşam ne kadar maddeyi kullanabilir ve ne kadar enerji, bilgi ve hesap çıkarabilir? Bu nihai sınırlar bizim kavrayışımız tarafından değil fizik kurallarınca koyulur. Bu da ironik biçimde, yaşamın geleceğini uzun vadede analiz etmeyi kısa vadeden daha kolay yapar.

Eğer 13.8 milyar yıllık kozmik tarihimiz bir haftaya sıkıştırılırsa, önceki iki bölümde anlatılan on bin yıllık drama yarım saniyeden az sürer. Bu da zekâ patlamasının nasıl ortaya çıkacağını ya da çıkıp çıkmayacağını ve sonrasında neler olacağını tahmin edemesek de tüm bu karmaşanın kozmik tarihte yaşamın nihai sınırlarını etkilemeyen bir flaş olacağı anlamına gelir. Patlama sonrası yaşam, bugünün insanlarının sınırları zorlamada olduğu kadar takıntılı olursa, teknoloji bu sınırlara *ulaşmak* üzere gelişecektir çünkü bunu yapabilecektir. Bu bölümde, bu sınırların ne olduğunu inceleyeceğiz ve uzun vadede yaşamın geleceğinin neye benzeyebileceğine dair bir içgörü kazanacağız. Bu sınırlar fizik hakkında şu anda sahip olduğumuz bilgilere dayandığı için, imkânların alt sınırı olarak görülmeleri gerekir: Gelecekteki bilimsel keşifler daha iyi fırsatlar sunabilir.

Gelecekteki hayatın gerçekten de çok hırslı olacağına emin miyiz? Hayır, değiliz: Bir eroin bağımlısı ya da sürekli *Keeping Up with the Kardashians* programının eski bölümlerini izleyen bir hımbıl kadar kayıtsız olabilir. Ancak hırsın yaşamın ilerlemesinde görece genetik bir özellik olduğundan şüphelenmek için

sebeplerimiz var. İster zekâ, ister yaşam süresi, bilgi ya da ilginç deneyimler olsun neyi maksimize ettiğinden neredeyse bağımsız olarak, kaynaklara ihtiyacı olacak. Bu yüzden de teknolojiyi nihai sınırlara itmek, sahip olduğu kaynaklardan mümkün mertebe faydalanmak için bir itkiye sahip olacaktır. Bundan sonra da daha da ilerlemenin tek yolu kozmosun daha büyük bölgelerine yayılarak daha fazla kaynak ele geçirmektir.

Ayrıca yaşam, bağımsız olarak kozmosumuzda pek çok yerde ortaya çıkmış olabilir. Bu durumda hırslı olmayan uygarlıklar kozmik olarak alakasız olurlar çünkü en hırslı yaşam formları kozmosu yavaş yavaş ele geçirir. Doğal seçilim de kozmik bir ölçekte rol oynar ve bir süre sonra kalan tüm yaşam yalnızca hırslı yaşam olur. Özet olarak, kozmosumuzun nihai olarak canlanabileceği noktayla ilgiliysek, fizik kurallarınca empoze edilen hırs sınırlarını araştırmalıyız. Hadi bunu yapalım! İlk olarak Güneş sistemimizde sahip olduğumuz kaynaklarla (madde, enerji *vs.*) neler yapılabileceğinin sınırlarını keşfedelim, sonra da kozmik keşif ve yerleşmeyle nasıl daha fazla kaynak bulabileceğimize bakalım.

Kaynaklarınızdan Mümkün Mertebe Faydalanmak

Bugünün süpermarketleri ve mal borsaları "kaynak" diyebileceğimiz on binlerce şey satarken, teknoloji sınırlarına ulaşmış gelecek yaşam temelde yalnızca tek bir ana kaynağa ihtiyaç duyacak: atomlar ve bileşenlerinden (kuarklar ve elektronlar) yapılmış her şeyi niteleyen ve *baryonik madde* adı verilen şey. Bu madde hangi formda olursa olsun, teknoloji bunu istenen madde ya da nesne haline getirebilecek şekilde yeniden düzenleyebilecek. Bunlar enerji santralleri ve bilgisayarlar da olabilir, ileri yaşam formları da. Bu yüzden de ileri düzey yaşama güç veren enerji-

nin ve bu yaşamın düşünmesini sağlayan bilgi işlemin sınırlarını irdeleyerek başlayalım.

Dyson Küreleri İnşa Etmek

Yaşamın geleceğinden bahsettiğimizde, en umutlu vizyonerlerden biri Freeman Dyson'dır. Son yirmi yıldır onu tanıdığım için onur ve mutluluk duyuyorum ancak onunla ilk tanıştığımda çok gergindim. Princeton'da İleri Çalışma Enstitüsü'nün yemek odasında arkadaşlarıyla bir şeyler atıştıran yeni bir post-doktora öğrencisiydim. Birden, eskiden Einstein ve Gödel'le takılan dünyaca ünlü bir fizikçi gelip kendini tanıtarak bize katılıp katılamayacağını sordu! Neyse ki öğle yemeğini eski kafalı yaşlı profesörlerle değil gençlerle yemeği tercih ettiğini söyleyerek beni hemen rahatlattı. Bu cümleleri yazarken doksan üç yaşında olmasına rağmen Freeman halen tanıdığım çoğu insandan ruhen daha gençtir ve gözlerindeki yaramaz çocuk parıltısı formaliteler, akademik hiyerarşiler ve genelgeçer bilgiyi hiç umursamadığını açık eder. Bir fikir ne kadar cesursa onun için o kadar heyecan vericidir.

Enerji kullanımından bahsettiğimizde, Sahra Çölü'nün %0,5'inden daha küçük bir alana düşen güneş ışığını toplayarak mevcut tüm küresel enerji ihtiyacımızı karşılayabileceğimizi söyleyip insanlar olarak çok hırssız olmamızla alay etmişti. Ama neden burada duralım ki? Dünyaya düşen tüm güneş enerjisini yakalamak bile yeterli değildi, neden Güneş'in enerji çıktısının *tümünü* alıp yaşamın hizmetine sunmuyorduk ki?

Ebeveyn yıldızının etrafında dönen yapay dünya halkalarının anlatıldığı Olaf Stapledon'un 1937 tarihli bilimkurgu klasiği *Star Maker*'dan ilham alarak, Freeman Dyson 1960 yılında sonradan *Dyson küresi* olarak anılacak şeyin tanımını yayımladı.[1] Freeman'ın fikri Jüpiter'i, gelecek nesillerimizin gelişeceği, insan-

lığın bugün kullandığından yüz milyar daha fazla biyokütle ve trilyon kat daha fazla enerjinin tadını çıkaracakları Güneş'i saran küresel bir kabuk şeklinde bir biyosfere dönüştürmekti.[2] Bunun doğal olarak gelecek bir sonraki adım olduğunu savunuyordu: "Endüstriyel gelişim aşamasına girdikten sonraki birkaç bin yıl içerisinde herhangi bir zeki türün ebeveyn yıldızını tamamen çevreleyen yapay bir biyosferi işgal etmesi beklenmelidir." Eğer bir Dyson küresinin içinde yaşarsanız hiç gece görmezsiniz: her zaman Güneş'i karşınızda bulursunuz ve tüm gökyüzünde, tıpkı şimdi güneş ışığının Ay'dan yansıdığını görmeniz gibi, biyosferin kalan kısmından yansıdığını görürdünüz. Eğer yıldızları görmek isterseniz, yalnızca "yukarı" gidip Dyson küresinin dışından kozmosa bir göz atmanız yeterli olur.

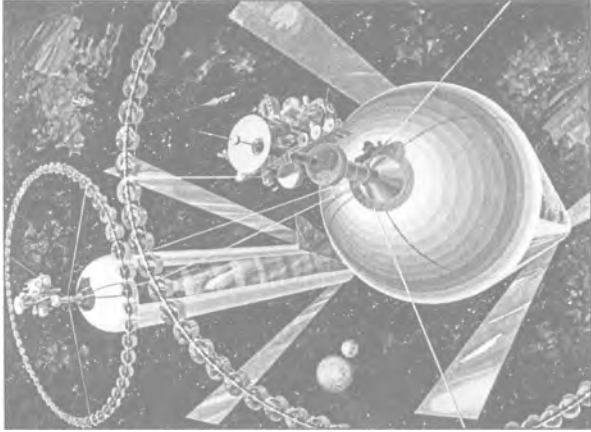
Kısmi bir Dyson küresi inşa etmenin düşük teknolojik yolu Güneş'in çevresine dairesel bir yörüngede bir halka boyunca yaşam ortamları yerleştirmektir. Güneşi tamamen çevrelemek için, çarpışmayı engellemek amacıyla, azıcık farklı uzaklıklarda farklı eksenlerde yörüngede dönen başka halkalar ekleyebilirsiniz. Bu hızlı dönen halkaların birbirine bağlı olmaması ve dolayısıyla ulaşım ve iletişimi zorlaştırması sıkıntısından kurtulmak için de bunun yerine Güneş'in içe doğru yerçekiminin radyasyonundan gelen dışa doğru itkiyle dengelendiği –Robert L. Forward ve Colin McInnes'in öncülüğünü yaptığı- yekpare bir sabit Dyson küresi inşa edebiliriz. Küre parça parça daha fazla "sadu" ekleyerek inşa edilebilir: Güneş'in yerçekimine merkezcil kuvvetten ziyade radyasyon basıncıyla karşı gelen sabit uydular. Bu iki kuvvet de Güneş'e uzaklığın karesiyle azalır. Bu da Güneş'ten bir uzaklıkta dengelenebiliyorlarsa, başka bir uzaklıkta da dengelenebilecekleri, Güneş sistemimizde herhangi bir yerde durabilecekleri anlamına gelir. Saduların metrekare başına yalnızca 0,77 gram olacak kadar hafif olmaları gerekir: yani kâğıttan yüz kat daha hafif. Fakat bu sorun değildir. Mesela, bir grafen levhası (kümes teline ben-

zer şekilde altıgen bir örüntüyle dizilmiş tek bir katman karbon atomları) bu limitten bin kat daha hafiftir. Eğer Dyson küresi güneş ışığının büyük çoğunluğunu emmekten ziyade yansıtacak şekilde inşa edilirse, içerisinde dolaşan ışığın toplam yoğunluğu ciddi oranda artarak radyasyon basıncını ve kürenin destekleyebileceği kütle miktarını da yükseltir. Pek çok başka yıldızın parlaklığı Güneş'imizden binlerce hatta milyonlarca kat fazladır ve bu yüzden de daha ağır sabit Dyson küreleri destekleyebilirler.

Eğer daha ağır katı Dyson kürelerine Güneş sistemimizde ihtiyaç duyulursa, Güneş'in yerçekimine karşı çıkmak için bükülmeden veya sıvılaşmadan, dünyanın en uzun gökdelenlerinin temelinde bulunanlardan on binlerce kat daha yüksek basınca dayanabilecek ultra-güçlü materyaller gerekir. Uzun ömürlü olması için bir Dyson küresinin dinamik ve akıllı olması, konumunu ve şeklini bozulmalara cevaben devamlı olarak düzenlemesi ve arada bir can sıkıcı asteroit ve kuyruklu yıldızların sorun yaşamadan geçebilmeleri için büyük delikler açabilmesi gerekir. Alternatif olarak, böyle sisteme izinsiz giriş yapan şeylerle baş etmek için tespit et ve saptır yöntemi de kullanabilir. Bu maddeler bileşenlerine ayrıştırılarak daha iyi amaçlar için kullanılabilir.

Bugünün insanları için Dyson küresinde ya da içinde yaşam en iyi ihtimalle kafa karıştırıcı, en kötü ihtimalle de imkânsız olur ancak bu geleceğin biyolojik ve biyolojik olmayan yaşam formlarının orada yaşayabilecek olmasını engellemez. Yörüngede dolaşan versiyonunda hiç yerçekimi yoktur ve sabit olanın yalnızca dış kısmında (Güneş'ten öteye bakan) alışkın olduğunuzdan on bin kat daha az yerçekimiyle düşmeden yürüyebilirsiniz. Sizi Güneş'ten gelen tehlikeli parçacıklardan koruyan bir manyetik alanınız (eğer inşa etmediyseniz) yoktur. İşin iyi yanı Dünya'nın şu anki yörüngesi büyüklüğünde bir Dyson küresi bize beş yüz milyon kat daha fazla yaşayacak alan sağlar.

Eğer daha çok Dünya benzeri yaşam alanları isteniyorsa, iyi haber bunların bir Dyson küresi inşa etmekten daha kolay olduğudur. Mesela, Görsel 6.1 ve 6.2 Amerikalı fizikçi Gerard K. O'Neill tarafından ortaya atılan, yapay yerçekimi, kozmik ışın kalkanı, yirmi dört saatlik gündüz gece döngüsü ve Dünya benzeri atmosfer ve ekosistemleri destekleyen silindirik bir yaşam alanı tasarımını gösterir. Böyle yaşam alanları bir Dyson küresi içerisinde özgürce dolaşabilir ya da modifiye edilmiş versiyonları kürenin dışına eklenebilir.

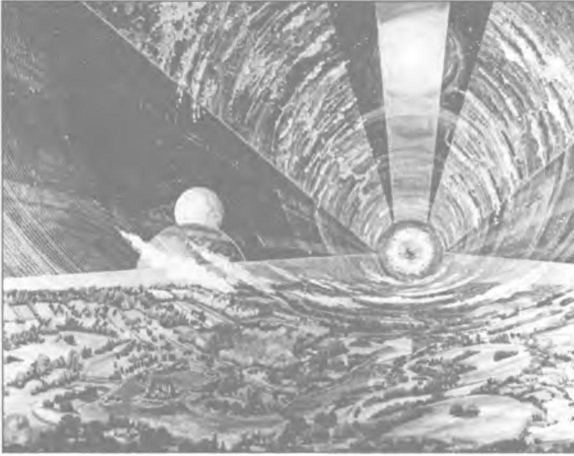


Görsel 6.1: Ters dönüşlü O'Neill silindir çifti. Daima Güneş'e bakacak şekilde onun etrafında dönerse Dünya benzeri, rahat insan yaşam alanları sağlanabilir. Rotasyonlarından ortaya çıkan merkezcil kuvvet suni yer çekimi sağlar ve üç katlanabilir ayna 24 saatlik gündüz gece döngüsü içerisinde güneş ışığını yansıtır. Halka şeklinde düzenlenmiş küçük alanlar tarım için özelleştirilmiştir. Görsel Rick Guidice/NASA izniyle kullanılmıştır.

Daha İyi Enerji Santralleri İnşa Etmek

Dyson küreleri bugünün mühendislik standartlarına göre verimli olsa bile, fizik kurallarının koyduğu sınırları zorlamanın yanına

bile yaklaşamaz. Kütleyi enerjiye %100 verimlilikle dönüştürürsek,* c ışık hızı olmak üzere, *m* miktarında kütlenin bize *E* miktarında enerji vereceğini ünlü formülü $E=mc^2$'yle Einstein belirtmektedir. Bu da *c* çok büyük olduğu için küçük miktar kütlenin bile devasa enerji üretebileceği anlamına gelir. Eğer yüksek miktarda anti madde kaynağımız olsa (ki yok), %100 verimli bir enerji santralini yapmak kolay olurdu. Bir çay kaşığı anti maddeyi suyun içine dökmek 200 bin ton TNT'ye denk enerji açığa çıkarırdı, yani bir hidrojen bombası kadar. Bu da yedi dakikalığına tüm dünyanın enerji ihtiyacını karşılardı.



Görsel 6.2: Bir önceki şekildeki O'Neill silindirlerinden birinin iç görünüşü. Eğer çapı 6,4 kilometre olursa ve her iki dakikada bir tur atarsa, yüzeyindeki insanlar Dünya'dakinin aynı yerçekimini deneyimlerler. Güneş arkanızdadır ancak geceleri kapanan, silindirin dışındaki bir aynadan dolayı yukarıda gözükür. Hava geçirmeyen pencereler atmosferin silindirden kaçmasını engeller. Görsel Rick Guidice/NASA izniyle kullanılmıştır.

* Eğer enerji sektöründe çalışıyorsanız, verimliliği, kullanılabilir bir biçimde salınmış enerji kısmı olarak tanımlamaya alışkın olabilirsiniz.



Görsel 6.3: İleri teknoloji, yiyerek ya da yakarak elde ettiğimize nazaran çok daha fazla enerjiyi maddeden çıkarabilir ve nükleer füzyon bile fizik kurallarının koyduğu sınırlardan 140 kat daha az enerji çıkarır. Sfaleronları, kuasarları ve buharlaşan kara delikleri kullanan enerji santralleri çok daha iyisini yapabilir.

Yöntem	**Verimlilik**
Çikolata sindirmek	%0,00000001
Kömür yakmak	%0,00000003
Benzin yakmak	%0,00000005
Uranyum-235 füzyonu	%0,08
Güneş ölene kadar Dyson küresi kullanmak	%0,08
Hidrojenin helyuma füzyonu	%0,7
Dönen kara delik motoru	%29
Kuasar çevresinde Dyson küresi	%42
Sfalerizer	%50?
Kara delik buharlaşması	%90

Tablo 6.1: Teorik sınır *E=mc²*'ye bağlı olarak kütleyi kullanılabilir enerjiye dönüştürme verimliliği. Metinde açıklandığı gibi, kara delikleri besleyerek ve buharlaşmalarını bekleyerek %90 verimlilik elde etmek kullanışlı olamayacak kadar yavaştır ve süreci hızlandırmak da verimliliği önemli ölçüde azaltır.

Öte yandan, bugün enerji üretme yollarımız, Tablo 6.1 ve Görsel 6.3'te gösterildiği üzere, berbat biçimde verimsizdir. Çikolatayı sindirmek içerdiği mc^2 enerjinin ancak on trilyonda birini elde ettiği için ancak %0,00000001 verimlidir. Eğer mideniz %0,001 verimli olsaydı, geri kalan hayatınız boyunca yalnızca tek bir öğün yemeniz yeterli olurdu. Yemeye kıyasla, kömür ve benzin yakmak ancak sırasıyla 3 ve 5 kat verimlidir. Bugünün nükleer reaktörleri uranyum atomlarını füzyon yoluyla ayırarak daha verimli işlerler ancak enerjilerinin %0,08'inden fazlasını çıkaramazlar. Güneşin çekirdeğindeki nükleer reaktör bizim inşa ettiklerimizden on kat daha verimlidir ve hidrojeni helyuma füzyon ederek enerjinin %0,7'sini çıkarır. Ancak Güneş'i mükemmel bir Dyson küresiyle sarsak bile, Güneş'in kütlesinin %0,08'inden daha fazlasını kullanılabilir enerjiye çeviremeyiz çünkü Güneş hidrojen yakıtının onda birini harcadığında, normal bir yıldız olarak yaşamını sonlandıracak, kırmızı deve dönüşecek ve ölmeye başlayacaktır. Diğer yıldızlar için de işler daha iyi değildir: Ana yaşamları süresince harcadıkları hidrojen çok küçük yıldızlardaki %4 orandan büyük yıldızlardaki %12'i orana kadar dar bir aralığa sahiptir. Eğer tüm hidrojenin %100'ünü füzyon edebilecek suni bir füzyon reaktörü yapabilirsek, yine de füzyon sürecinde %0,7 gibi utanç verici düşüklükte bir verimlilikte kalırız. Daha iyisini nasıl yapabiliriz?

Buharlaşan Kara Delikler

Zamanın Kısa Tarihi isimli kitabında Stephen Hawking bir kara delik enerji santrali önerisinde bulunur.* Kara deliklerin uzun zamandır ışığın bile kaçamadığı tuzaklar olduğu düşünüldüğünde

* Eğer yakınlarda hiçbir uygun doğal kara delik bulunamazsa, yeterince küçük bir alana çok fazla madde koyularak yeni bir tanesi de yaratılabilir.

bu paradoksal görünür. Ancak bilindiği üzere Hawking, kuantum yerçekiminin kara deliklerin artık *Hawking radyasyonu* adı verilen ısı radyasyonu yayan sıcak bir obje gibi –ne kadar küçükse o kadar sıcak– davranmasını sağladığını hesaplamıştı. Bu da kara deliklerin zaman içinde enerjilerini kaybedip yok olduğu anlamına gelir. Diğer bir deyişle, kara deliğe hangi madde girerse girsin ısı radyasyonu olarak geri gelecektir. Yani kara delik tamamen yok olduğunda, maddenizi radyasyona yaklaşık %100 verimlilikle çevirmiş olursunuz.*

Kara delik buharlaşmasını bir enerji kaynağı olarak kullanmaktaki bir problem de kara delik bir atomdan çok daha küçük değilse eğer, bu sürecin evrenimizin şimdiki yaşından daha da uzun süren ve bir mumdan daha az enerji yayan acı verici yavaşlıkta bir süreç olmasıdır. Üretilen enerji deliğin boyutunun karesi oranıyla azalır ve fizikçiler Louis Crane ve Shawn Westmoreland de gelmiş geçmiş en büyük deniz aracı kadar ağır ama bir protondan bin kat daha küçük bir kara delik kullanmayı teklif etmişlerdi.[3] Ana motivasyonları da kara delik motorunu bir uzay gemisine güç sağlamak için kullanmaktı (bu konuya ileride yine döneceğiz), bu yüzden de verimlilikten çok taşınırlığı önemsiyorlardı ve kara deliği lazer ışığıyla beslemeyi önerdiler, bu da hiçbir enerji-madde dönüşümünün olmamasına sebep oldu. Radyasyon yerine maddeyle beslemiş olsalardı bile yüksek verimliliği garanti altına almak zor olurdu: Binde birleri kadar küçük bir kara deliğe protonların girmesi için bu deliğe en az Büyük Hadron Çarpıştırıcısı kadar

* Bu biraz fazla basite indirgemedir çünkü Hawking radyasyonu, ayrıca faydalı iş çıkarmanın zor olduğu bazı parçacıkları da içerir. Büyük kara delikler ancak yüzde 90 verimlidir çünkü enerjinin yüzde 10'u graviton formunda dışa verilmiştir. Yani, değil faydalı iş olarak kullanılmaları, tespit edilmeleri bile neredeyse imkânsız olan aşırı utangaç parçacıklar. Kara delik buharlaşmaya ve küçülmeye devam ederken, verimlilik daha da düşer çünkü Hawking radyasyonu nötrino ve diğer büyük parçacıkları da dâhil etmeye başlar.

güçlü bir makine tarafından fırlatılmalı, mc^2 enerjileri en az bin kat daha fazla kinetik (hareket) enerjisiyle yükseltilmeliydi. Kara delik buharlaştığında bu kinetik enerjinin en az %10'u gravitonlara kaybedileceği için, çıkarabileceğimiz ve işe koşabileceğimizden daha fazla enerji kara deliğe verilmiş, negatif verimlilik elde edilmiş olunacaktı. Kara delik enerji santralinin ihtimallerini daha da karıştıran şey hesaplamalarımızı üzerinden yapabileceğimiz güçlü bir kuantum yerçekimi teorisine hâlâ sahip olmamamızdır ancak bu belirsizlik elbette daha keşfedilecek yeni ve faydalı kuantum yerçekimi etkileri olduğu anlamına da gelebilir.

Dönen Kara Delikler

Neyse ki kuantum yerçekimi ya da diğer daha pek anlaşılmamış fizik yasalarını dâhil etmeden kara delikleri enerji santralleri olarak kullanmanın başka yolları da vardır. Mesela, pek çok kara delik çok hızlı döner, olay ufukları da neredeyse ışık hızında döner ve bu rotasyon enerjisi elde edilebilir. Bir kara deliğin olay ufku ışığın bile kaçamadığı bölgedir çünkü yerçekimi kuvveti çok güçlüdür. Görsel 6.4 olay ufkunun dışında, dönen bir kara deliğin boşluğu da kendisiyle birlikte bir parçacığın durmasının ya da sürüklenmemesinin imkânsız olacağı hızda sürüklediği *ergosfer* adı verilen bir bölge olduğunu gösterir. Eğer ergosfere bir nesne atarsanız, deliğin etrafında dönerek hızlanır. Maalesef, kısa süre içinde kara delik tarafından yutularak olay ufkunda sonsuza dek yok olacaktır. Bu yüzden de enerji elde etmek istiyorsanız hiçbir işinize yaramayacaktır. Ancak Roger Penrose zekice bir açıyla bir nesneyi fırlatır ve Görsel 6.4'te gösterildiği gibi iki parçaya bölünmesini sağlarsanız, tek bir parçanın yutulurken diğerinin kara delikten başlangıçtakinden daha yüksek enerjiyle kaçabilmesini sağlayabileceğinizi göstermiştir. Diğer bir deyişle, kara deliğin dönüş enerjisinin bir kısmını işe koşabileceğiniz faydalı enerjiye

başarılı bir biçimde dönüştürmüş olursunuz. Bu süreci pek çok kereler tekrar ederek, kara deliğin dönmesi durana ve ergosferi yok olana kadar *tüm* dönüş enerjisinden faydalanabilirsiniz. Eğer başta kara delik doğanın izin verdiği kadar hızlı dönüyorsa ve olay ufku da ışık hızındaysa, bu strateji size kütlesinin %29'unu enerjiye dönüştürme imkânı verir. Hâlâ kara deliklerin ne kadar hızlı döndüğüne dair ciddi bir belirsizlik vardır ancak üzerinde en çok araştırma yapılanların pek çoğu oldukça hızlı dönüyor gibi duruyor: maksimumun %30'u ila %100'ü arasında. Galaksimizin ortasındaki devasa kara delik (ağırlığı Güneş'in dört milyon kat fazlasıdır) de dönmektedir, bu yüzden de eğer onun kütlesinin %10'u bile yararlı enerjiye dönüştürülebilirse, %100 verimlilikle çevrilmiş dört yüz bin güneşin enerjisine ya da milyarlarca yıl boyunca beş yüz milyon Güneş'ten Dyson küreleriyle elde edeceğimiz toplam enerjiye denk olacaktır.



Görsel 6.4: Kara deliğe yakın bir yere bir A parçacığı atarak ve onun yutulan C parçası ve kaçan B parçası şeklinde ikiye ayrılmasını sağlayarak kara deliğin dönüş enerjisinin bir kısmı elde edilebilir, hem de başta A'nın sahip olduğundan daha fazla enerjiyle.

Kuasarlar (Yıldızsılar)

Bir diğer ilginç strateji kara delikten değil de içine düşen maddeden enerji çıkarmaktır. Doğa bunu kendi kendine yapmanın bir yolunu çoktan bulmuştur: kuasar. Gazlar bir kara deliğe yakınlaşırken girdap halini alır. En iç kısımları yutulan, pizza biçimli bir diske dönüşürken aşırı derecede ısınır ve bol miktarda radyasyonu dışarı verir. Gaz deliğe doğru düşmeye başlarken hızlanır ve tıpkı paraşütle atlayan bir hava dalışçısının yapacağı gibi yerçekimsel potansiyel enerjiyi hareket enerjisine dönüştürür. Hareket giderek daha da dağınık bir hal alırken karmaşık türbülans, gaz kütlesinin koordine hareketini giderek daha da küçülen ölçeklerde rastgele harekete çevirir. Ta ki tekil atomlar yüksek hızlarda birbirlerine çarpmaya başlayana dek. Böylesine rastgele harekete sahip olmak tam olarak sıcak olmaya denktir ve bu sert çarpışmalar hareket enerjisini radyasyona dönüştürür. Tüm kara deliğin etrafına, güvenli bir mesafede, bir Dyson küresi inşa ederek bu radyasyon enerjisi yakalanabilir ve kullanılabilir. Kara delik ne kadar hızlı dönerse süreç o kadar verimli olur. En hızlı seviyede dönen bir kara delik %42 gibi ciddi bir verimlilik oranı gösterebilir.* Yaklaşık bir yıldız kadar ağır olan kara delikler için, çoğu enerji X-ray olarak çıkarken galaksilerin merkezinde bulunan süper kütleli cinsleri için, çoğu kızılötesi, görülür ve ultraviyole ışık olarak ortaya çıkar.

Kara deliğinizi besleyecek yakıtınız bittiğinde, yukarıda da bahsettiğimiz gibi dönüş enerjisini elde etmeye başlayabilirsiniz.**

* Douglas Adams hayranları için, bunun yaşam, evren ve her şeye dair soruya cevap veren zekice bir soru olduğunu dikkate alın. Daha da net olmak gerekirse verimlilik $1-1/\sqrt{3} \approx \%42$'dir.

** Eğer aynı yönde yavaşça dönen bir gaz bulutunu kara deliğin etrafına koyarak onu beslerseniz, bu gaz içeri çekilip yutulurken daha da hızlı dönecek ve tıpkı bir artistik patinajcının kollarını içeri çektiğinde daha hızlı dönmesi gibi kara deliğin dönüşünü de artıracaktır. Bu da deliği maksimum seviyede döner biçimde tutarak önce gaz enerjisinin %42'sini, sonra da kalanın %29'unu çıkarmanızı sağlar ve toplam verimlilik $\%42 + (1-\%42) \times \%29 \approx \%59$ olur.

Gerçekten de doğa, Blandford-Znajek mekanizması olarak bilinen bir manyetik süreç yoluyla artan gazdan gelen radyasyonu artırarak bunu kısmen yapmanın bir yolunu çoktan bulmuştur. Manyetik alanların veya diğer bileşenlerin zekice kullanımıyla, teknolojiyi enerji çıkarma verimliliğini %42'nin daha da ilerisine taşımakta kullanması da gayet mümkündür.

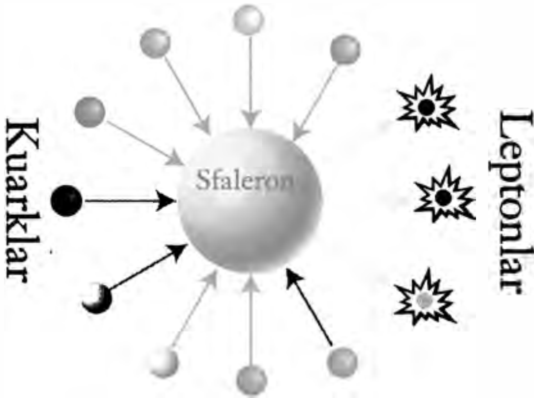
Sfaleronlar

Maddeyi enerjiye çevirmenin kara delikleri içermeyen başka bir bilinen yolu da vardır: *sfaleron* süreci. Bu süreç, kuarkları yok edebilir ve onları leptonlara, yani elektronlara, daha ağır kuzenleri muon ve tau parçacıklarına, nötrinolara veya antipartiküllerine çevirebilir.[4] Görsel 6.5'te de gösterildiği üzere, parçacık fiziğinin standart modeli uygun nitelik ve dönüşe sahip dokuz kuarkın bir araya gelebileceğini ve sfaleron adı verilen bir ara durum yoluyla üç leptona dönüşebileceğini tahmin eder. Girdi çıktıdan daha ağır olduğu için, kütle farkı Einstein'ın $E=mc^2$ formülüne göre enerjiye dönüşür.

Geleceğin zeki yaşamı da benim *sfalerizer* adı verdiğim şeyi yapmayı başarabilir: abartılı bir dizel motoru gibi çalışan bir enerji jeneratörü. Geleneksel bir dizel motoru kendiliğinden alev alıp yanmaları için sıcaklık yeterince yüksek olana dek hava ve motorin karışımını sıkıştırır, sonrasında da sıcak karışım yeniden genleşir ve pistonu itmek gibi faydalı bir iş yapar. Karbondioksit ve diğer yanma gazları başlangıçta pistonda olandan %0,00000005 kadar daha hafiftir ve bu kütle farkı motoru çalıştıran ısı enerjisine dönüşür. Sfalerizer sıradan maddeyi birkaç katrilyon dereceye kadar sıkıştıracak ve sfaleronlar işlerini yaptıktan sonra yeniden genleşip soğumasına izin verecektir.* Bu deneyin sonucunu çoktan

* Elektromanyetik ve zayıf kuvvetleri yeniden birleştirmek için yeterince sıcak olması gerekir. Bu da parçacıklar, bir parçacık çarpıştırıcısında iki yüz milyar volt ile hızlandırıldıkları kadar hızlı hareket ettiklerinde olur.

biliyoruz çünkü evrenimizin erken dönemi bize bunu sıcak olduğu, yaklaşık 13.8 milyar yıl önce de göstermişti: Maddenin neredeyse %100'ü enerjiye dönüşür, parçacıkların yalnızca milyarda biri kalır ve onlar da sıradan maddeyi oluşturan temel parçacıklardır: kuarklar ve elektronlar. Yani gerçekten de bir dizel motoru gibidir, sadece bir milyar kattan daha fazla verimlidir! Diğer bir avantajı ise neyle besleyeceğiniz konusunda titiz olmak zorunda değilsiniz – kuarklardan oluşmuş her şeyle çalışır, yani herhangi bir normal maddeyle.



Görsel 6.5: Parçacık fiziğinin standart modeline göre, uygun nitelik ve dönüşe sahip dokuz kuark bir araya gelir ve sfaleron adı verilen bir ara durumun ardından üç leptona dönüşebilir. Kuarkların toplam kütlesi (onlara eşlik eden gluon parçacıklarının enerjisiyle birlikte) leptonların kütlesinden çok daha fazladır; bu yüzden de bu süreç şekildeki patlamaların gösterdiği üzere enerji salar.

Bu yüksek sıcaklıklı süreçlerden ötürü bebek evrenimiz maddeden (sonradan atomlarla birleşen kuarklar ve elektronlar) bir trilyondan fazla kat daha fazla radyasyon (fotonlar ve nötrinolar) üretir. 13.8 milyar yıldan beri büyük bir ayrım yaşanmıştır. Atomlar galaksiler, yıldızlar ve gezegenler halinde yoğunlaşırken, çoğu

foton da galaksiler arası boşlukta kalarak evrenimizin bebeklik fotoğraflarını çekebilmemizi sağlayan kozmik mikrodalga arka plan radyasyonunu oluşturur. Bir galakside ya da diğer madde konsantrasyonlarında yaşayan herhangi bir ileri yaşam formu mevcut maddesinin büyük kısmını enerjiye dönüştürebilir ve sfalerizerin içinde kısa süre için o sıcak yoğun koşulları yeniden yaratarak erken dönem evrenimizden ortaya çıkan aynı küçük değere madde oranını düşürür.

Gerçek bir sfalerizerin ne kadar verimli olacağını bulabilmek için ana pratik detayların üzerinde çalışmamız gerekir: Mesela, sıkıştırma aşamasında ciddi oranda foton ve nötrino kaybını önlemek için ne kadar büyük olmalıdır? Kesin olarak söyleyebileceğimiz bir şey var ise gelecekteki yaşamın enerji ihtimallerinin şimdiki teknolojimizin izin verdiğinden çok daha iyi olacağıdır. Bir füzyon reaktörü inşa etmeyi bile başaramadık ancak gelecek teknolojisi on hatta yüz kat daha iyisini yapmayı başaracaktır.

Çok Daha İyi Bilgisayarlar Yapmak

Eğer akşam yemeği yemek enerji verimliliğinin fiziksel limitinden on milyar kat kötüyse, bugünün bilgisayarları ne kadar verimlidir? Birazdan göreceğimiz üzere akşam yemeğinden çok daha kötüdür.

Sık sık dostum ve meslektaşım Seth Lloyd'u MIT'de benim kadar deli olduğu söylenebilecek tek kişi olarak tanıştırırım. Kuantum bilgisayarları hakkında öncü çalışmalar yaptıktan sonra, tüm evrenimizin bir kuantum bilgisayarı olduğunu iddia ettiği bir kitap da yazdı. Sık sık işten sonra bira içeriz ve henüz söyleyecek ilginç bir şeye sahip olmadığı bir konu olduğunu görmedim. Mesela, Bölüm 2'de söylediğim gibi hesaplamanın nihai sınırları

hakkında da söyleyecek çok şeyi vardır. 2000 yılında yayımladığı ünlü bir makalede, hesaplama hızının enerjiyle sınırlandırıldığını göstermişti: T süresinde temel bir mantık işlemini gerçekleştirmek için, h'nin Planck sabiti olarak da bilinen temel fizik niceliği olduğu durumda, ortalama $E=h/4T$ kadar enerji gerekmektedir. Bu da bir kiloluk bilgisayarın saniyede en çok 5×10^{50} operasyon gerçekleştirebileceği anlamına gelir. Yani bu cümleleri yazdığım bilgisayardan 10^{36} kat fazlası. Hesaplama gücümüz her birkaç yılda bir iki katına çıkmayı sürdürürse birkaç yüzyıl içinde buraya geleceğimizi Bölüm 2'de görmüştük. Ayrıca, bir kiloluk bilgisayarın benim laptopumdan milyar milyar kat daha fazla, 10^{31} bit depolayabileceğini de göstermiştir.

Seth bu sınırlara erişmenin süper zeki yaşam için bile zor olacağını çünkü bir kiloluk nihai "bilgisayarın" termonükleer bir patlamaya ya da Büyük Patlama'nın küçük bir parçasına benzeyeceğini itiraf eden ilk kişiydi. Yine de pratik sınırların nihai olanlardan çok da uzak olmadığı konusunda da iyimserdir. Gerçekten de mevcut kuantum bilgisayar prototipleri de her atomda bir bit saklayarak hafızalarını olabildiğince minyatürleştirmişlerdir ve bunu büyütmek kilogram başına 10^{25} bit depolamaya izin verecek; yani laptopumdan trilyon kat daha iyi. Dahası, bu atomlar arasında iletişim kurmak için elektromanyetik radyasyonu kullanmak saniye başına 5×10^{40} operasyonu mümkün kılar; benim CPU'mdan 10^{31} kat fazla.

Özetle, gelecek yaşamın hesaplama ve keşfetme potansiyeli tamamıyla akıllara durgunluk verecek cinsten: Bugünün en iyi süperbilgisayarları nihai bir kiloluk bilgisayara, yalnızca tek bir bit bilgi depolayan, saniyede bir kere açılıp kapanan arabadaki dönüş sinyaline olduğundan daha uzaktır.

Diğer Kaynaklar

Fizik açısından, gelecek yaşamın yaratmak istediği her şey –yaşam alanlarından makinelere ve yeni yaşam formlarına– belirli bir şekilde düzenlenmiş temel parçacıklardan başka bir şey değildir. Tıpkı mavi bir balinanın yeniden düzenlenmiş karides, karidesin de yeniden düzenlenmiş plankton olması gibi tüm Güneş sistemimiz 13.8 milyar yıllık kozmik evrim boyunca yeniden düzenlenmiş hidrojendir: Yerçekimi hidrojeni yeniden yıldızlara, yıldızlar da hidrojeni daha ağır atomlara dönüştürür. Ardından yerçekimi böyle atomları gezegenimiz olarak yeniden düzenler ve buradaki kimyasal ve biyolojik süreçler onları yaşama dönüştürür.

Teknolojik sınırlarına ulaşan gelecek yaşam, böyle parçacık yeniden düzenlemelerini ilk önce hesaplama gücüyle en etkili yöntemi tespit ederek, sonrasında da mevcut enerjisini madde yeniden düzenleme sürecine güç sağlayarak daha hızlı ve verimli yapar. Maddenin nasıl hem bilgisayar hem de enerjiye dönüşebileceğini gördük, bu yüzden de bir bakıma ihtiyaç duyulan tek temel kaynaktır.* Gelecek yaşam, maddeyle yapabileceklerinin fiziksel sınırlarına geldiğinde daha fazlasını yapabilmek için elinde kalan tek bir yöntem olacak: Daha fazla madde elde etmek. Bunu yapmanın tek yolu da evrenimize doğru genişlemektir. Uzaya doğru ileri!

Kozmik Yerleşim Yoluyla Kaynak Elde Etmek

Kozmik birikimimiz ne kadar büyüktür? Özellikle de yaşamın faydalanacağı miktarda maddeye fizik kuralları hangi üst sınırı koyar? Kozmik birikimimiz akıllara durgunluk verecek kadar

* Yukarıda yalnızca atomlardan yapılmış maddeyi tartıştık. Yaklaşık sekiz kat daha fazla karanlık madde vardır ancak ele geçmez ve yakalaması zordur, rutin bir biçimde dünyanın içinden öte tarafa geçer, bu yüzden, gelecek yaşamın onu yakalayıp kullanıma sokup sokamayacağını göreceğiz.

büyüktür elbette ama tam olarak ne kadar? Tablo 6.2 bazı ana rakamları listelemektedir. Maddesinin biyosferimize dâhil olmayan ve yerçekimsel kuvvet ve manyetik alan sağlamak dışında yaşam için hiçbir faydası olmayan kısmı açısından gezegenimiz %99,999999 ölüdür. Bu da yaşama aktif destek verecek yüz milyon kat fazla maddeyi bir gün kullanabilme potansiyelini sunar. Eğer Güneş sistemimizdeki (Güneş de dâhil) tüm maddeyi optimal kullanıma sunabilirsek, bir milyon kat daha iyi olur. Galaksimizin her köşesine yerleşmek de kaynaklarımızı trilyon kat büyütecektir.

Bölge	Parçacık
Biyosferimiz	10^{43}
Gezegenimiz	10^{51}
Güneş sistemimiz	10^{57}
Galaksimiz	10^{69}
Işık hızının yarısında seyahat edersek menzilimiz	10^{75}
Işık hızında seyahat edersek menzilimiz	10^{76}
Evrenimiz	10^{78}

Tablo 6.2: Gelecek yaşamın kullanabilme ihtimali olan madde parçacıklarının (protonlar ve nötronlar) yaklaşık sayısı.

Ne Kadar Uzağa Gidebilirsiniz?

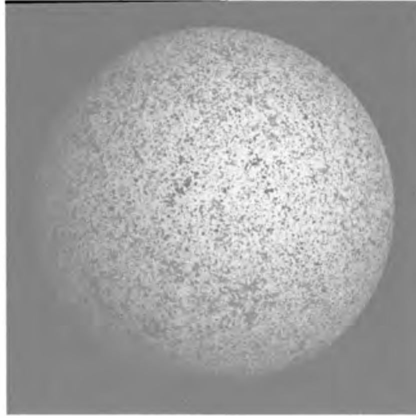
Eğer yeterince sabırlıysak istediğimiz kadar çok galaksiye yerleşerek sınırsız kaynak elde edebiliriz diye düşünüyor olabilirsiniz ancak modern kozmoloji böyle demiyor! Evet, uzayın kendisi sonsuz olabilir, sınırsız sayıda galaksi, yıldız ve gezegen içeriyor olabilir; gerçekten de 13.8 milyar yıl önce Büyük Patlama'yı neyin oluşturduğuna dair bugünkü en popüler bilimsel paradigma olan *enflasyonun* en basit versiyonlarının tahmin ettiği de budur. Ancak sonsuz galaksiler olsa da yalnızca sonlu sayıda kesimini görebilir ve erişebiliriz: Yaklaşık iki yüz milyar galaksi görebiliyoruz ve en çok on milyarına yerleşebiliriz.

Bizi kısıtlayan şey ışık hızıdır: yılda sadece bir ışık yılı (yaklaşık on trilyon kilometre). Görsel 6.6 uzayın Büyük Patlama'dan bu yana geçen 13.8 milyar yıl içerisinde bize ışığının ulaştığı bir kısmını, "gözlemlenebilir evrenimiz" ya da sadece "*bizim evrenimiz*" diye bilinen küresel bölgeyi göstermektedir. Eğer uzay sonsuzsa bile, bizim evrenimiz sonludur ve "yalnızca" 10^{78} atom içerir. Dahası, evrenimizim yaklaşık %98'i "görülebilir ama dokunulamazdır." Yani görebileceğimiz ancak ışık hızında yolculuk etsek bile sonsuza dek ulaşamayacağımız yerlerdir. Peki neden? Sonuçta, görebileceğimiz yerin uzaklık sınırları da evrenimizin sonsuz yaşlı olmadığı gerçeğine dayanır. Bu yüzden de uzak ışığın henüz bize varacak zamanı olmamıştır. O zaman yolda ne kadar zaman harcayacağımız önemli değilse, rastgele uzak galaksilere gidemez miyiz?

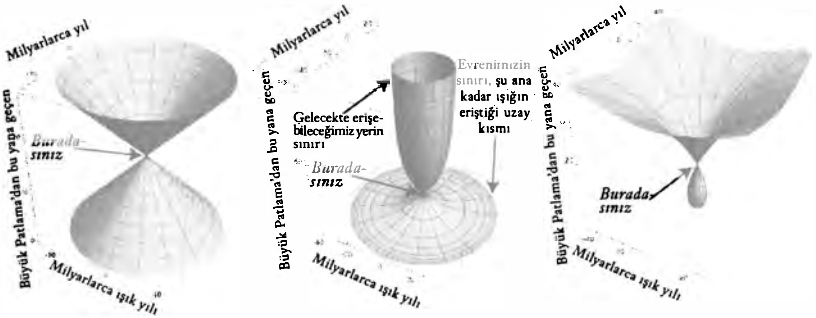
İlk zorluk evrenimizin genişlemesidir. Yani tüm o galaksiler bizden uzaklaşmaktadırlar, bu yüzden de uzak galaksilere yerleşmek bir kovalamaca oyununa dönüşür. İkinci zorluk evrenimizin %70'ini oluşturan gizemli kara enerjiden ötürü kozmik genleşmenin hızlanmasıdır. Bunun nasıl sorun yarattığını anlamak için, bir tren platformuna girdiğinizi ve treninizin yavaşça hızlanışını gördüğünüzü ama davetkâr bir biçimde bir kapının açık bırakıldığını düşünün. Eğer hızlı ve gözüpekseniz, treni yakalayabilir misiniz? Sonuçta koşabileceğinizden daha hızlı gideceği için, cevap tamamen başta trene olan uzaklığınıza bağlıdır: eğer belirli bir kritik uzaklıktan uzaktaysa onu asla yakalayamazsınız. Aynı durum, bizden sürekli hızlanarak uzaklaşan o uzak galaksileri yakalamak konusunda da geçerlidir: Işık hızıyla seyahat etsek bile, on yedi milyar ışık yılından daha uzakta bulunan galaksilere asla ulaşamayız – bu da evrenimizdeki galaksilerin %98'inden fazlasıdır.

Ama bir dakika durun: Einstein'ın özel görelilik teorisi hiçbir şeyin ışık hızından daha hızlı hareket edemeyeceğini söylemiyor muydu? O zaman galaksiler ışık hızında yolculuk eden bir şeyi nasıl geride bırakabilir? Cevap özel görelilik teorisinin hız sınırının daha bağımsız olduğu Einstein'ın genel görelilik teorisi tarafından yerinin alınmasındadır: Hiçbir şey *boşlukta* ışık hızından daha hızlı hareket edemez ancak boşluk istediği kadar hızlı genişlemekte özgürdür. Einstein ayrıca zamanı *uzayzamanda* dördüncü bir boyut olarak görerek bize bu hız sınırlarını canlandırmanın güzel bir yolunu da vermiştir (bkz. Görsel 6.7, burada üç uzay boyutundan birini kaldırarak her şeyi üç boyutlu tuttum). Eğer uzay genişlemeseydi, ışık ışınları uzayzaman içinde 45 derece eğimli doğrultular kurardı, böylece görebildiğimiz ve buradan ulaşabileceğimiz bölgeler koni olurdu. Geçmiş ışık konimizin yalnızca 13.8 milyar yıl önce Büyük Patlama'yla ucu kesilmişken, gelecekteki ışık konimiz sonsuza dek genişler, bize sınırsız bir kozmik birikime erişim olanağı verirdi. Aksine, şeklin orta paneli karanlık maddeyle genişleyen evrenin ışık konilerimizi şampanya bardağı şekline deforme edip yerleşebileceğimiz galaksi sayısını sonsuza dek on milyona indirdiğini göstermektedir.

Eğer bu sınır size kozmik klostrofobi veriyorsa, muhtemel bir kaçamakla sizi keyiflendirmeme müsaade edin: Yaptığım hesaplar, son saptanan bulgularla uyumlu bir biçimde karanlık enerjinin zamanla değişmediğini varsayıyor. Ancak karanlık enerjinin gerçekte ne olduğuna dair hiçbir fikrimiz yok. Bu da karanlık enerjinin giderek kaybolacağına dair bir umut verir (tıpkı kozmik enflasyonu açıklayan karanlık enerji benzeri madde gibi) ve bu olursa, hızlanma biter ve *yavaşlama* başlar, gelecek yaşam formlarının yaşamlarını sürdürdükçe yeni galaksilere yerleşebilmelerini potansiyel olarak mümkün kılabilecektir.



Görsel 6.6: Evrenimiz, yani ışığının Büyük Patlama'dan bu yana geçen 13.8 milyar yılda bize (merkezde) ulaşacak zamanı olan küresel uzay bölgesi. Bu örüntü evrenimizin Planck uydusu tarafından çekilmiş bebeklik fotoğraflarını, yani ancak dört yüz bin yaşında olduğu dönemi gösteriyor. Neredeyse Güneş'in yüzeyi kadar sıcak plazmadan oluşmaktadır. Uzay, muhtemelen, bu bölgenin ötesinde devam etmektedir ve her yıl görünür alana yeni madde gelmektedir.



Görsel 6.7: Bir uzayzaman diyagramında, bir olay, yatay ve dikey konumlarının sırasıyla nerede ve ne zaman gerçekleştiğini söylediği bir noktadır. Eğer uzay genişlemiyorsa (sol panel) iki koni Dünya'da (tepe noktasında) etkilenebileceğimiz (alt koni) ve etkimizin olacağı (üst koni) uzayzaman kısımlarını kapsar. Çünkü nedensel etkiler yılda bir ışık yılı kadar mesafe katedebilen ışıktan daha hızlı gidemez. Uzay genişlediğinde her şey daha ilginç bir hal alır (sağ panel). Standart kozmoloji modeline göre, uzay son-

suz olsa da uzayın yalnızca sonlu bir kısmını görebilir ve erişebiliriz. Bir şampanya bardağını anımsatan ortadaki imajda, uzak galaksilerin zamana göre hareketleri dikey çizgilere denk gelecek şekilde uzayın genişlemesini gizleyecek koordinatlar kullandık. Mevcut görüş açımızdan, Büyük Patlama'dan 13.8 milyar yıl sonra, ışık bize yalnızca şampanya bardağının altından erişebilmiştir ve ışık hızında hareket edebiliyor olsak bile, asla bardağın on milyar galaksiyi içeren üst kısmının dışındaki bölgelere ulaşamayız. Bir çiçeğin altındaki su damlacığına benzeyen sağdaki imajda, uzayın genişlemesinin görüldüğü bildiğimiz koordinatları kullandık. Bu, bardağın tabanını bir damlacık şekline dönüştürür çünkü görebildiğimiz yerin sınırlarındaki bölgeler başlangıçta birbirine çok yakındır.

En Çok Ne Kadar Hızlı Gidebilirsiniz?

Yukarıda ışık hızında tüm yönlere genişlendiğinde bir uygarlığın kaç tane galaksiye yerleşebileceğini soruşturduk. Genel görelilik uzaya ışık hızında roket göndermenin imkânsız olacağını çünkü bunun sonsuz enerji gerektireceğini söyler, o zaman roketler pratikte ne kadar hızlı gidebilir?*

NASA'nın New Horizons roketi 2006 yılında Plüton'a giderken saatte yüz bin mil (saniyede kırk beş kilometre) yaparak hız rekorunu kırdı ve NASA'nın 2018 Solar Probe Plus'ı Güneş'e çok yaklaşarak bunu dört kat hızlı yapmayı planlıyor ancak bu bile ışık hızının %0,1'inden daha azdır. Daha hızlı ve daha iyi roketleri bulmak için yapılan araştırma geçmiş yüzyılda bazı en parlak beyinlerin dikkatini çekmiştir ve bu konu üzerine ilgi çekici ve zengin bir literatür bulunmaktadır. Hızlı gitmek neden çok zordur? İki ana problem, konvansiyonel roketlerin yakıtlarının çoğunu taşıdıkları yakıtı hızlandırmakta kullanması ve günümüz roket yakıtının çaresizce verimsiz olmasıdır – enerjiye dönüş-

* Kozmik matematik olabildiğince basittir: Eğer uygarlık genişleyen uzaya doğru ışık hızı c ile değil daha düşük bir hız olan v ile genişlerse, yerleşilebilecek galaksi sayısı $(v/c)^3$ oranında azalır. Bu da hantal uygarlıkların ciddi anlamda cezalandırılacağı anlamına gelir. 10 kat yavaş başlayanın 1000 kat az galaksisi olur.

türülen kütle oranı Tablo 6.1'de gördüğümüz %0,00000005'lik benzinden daha iyi değildir. Çok açık bir iyileştirme de daha verimli bir yakıta geçmek olacaktır. Örneğin, Freeman Dyson ve başkaları, NASA'nın bir yüzyıl sürecek yolculukla insanları başka bir güneş sistemine taşımaya yetecek kadar büyük bir uzay gemisinin ışık hızının %3'üne ulaşabilmesi için on gün boyunca üç yüz bin kadar nükleer bomba patlatmayı planlayan projesi Orion'da çalışmışlardı.[5] Diğerleri de anti maddeyi yakıt olarak kullanmayı keşfetmişlerdi çünkü anti maddeyi normal maddeyle birleştirmek, neredeyse %100 verimli biçimde enerji salardı.

Bir diğer popüler fikir de kendi yakıtını taşımak zorunda olmayan bir roket inşa etmektedir. Mesela, yıldızlararası uzay mükemmel vakumda değildir ve hidrojen iyonları içerir (yalnız bir proton: elektronunu kaybetmiş bir hidrojen atomu). 1960'ta bu, fizikçi Robert Bussard'a *Bussard ramjet* olarak bilinen şeyin ardındaki fikri verdi: Böyle iyonları yol üzerinden çekecek ve aracın üzerindeki füzyon reaktöründe onları roket yakıtı olarak kullanacak. Son zamanlarda yapılan çalışmalar bunun pratikte yapılabilirliğine gölge düşürmüş olsa da yüksek teknolojili, uzay yolculuğu yapan bir uygarlığın yapabileceği başka bir yakıt taşımama fikri vardır: lazer yelkenciliği.

Görsel 6.8, Dyson küresi inşaatından bahsederken andığımız saduları da icat eden fizikçi Robert Forward tarafından 1984'te ortaya atılmış zekice bir lazer yelken roket tasarımını sergilemektedir. Bir yelkenlinin yelkeninden seken hava moleküllerinin onu itmesi gibi, aynadan seken ışık parçacıkları (foton) da bu roketi itecektir. Bir uzay aracına bağlı olarak geniş ve ultra hafif bir yelkene devasa güneş ışığı destekli lazer göndererek, Güneş'in enerjisini roketi yüksek hızlara çıkarmak için kullanabiliriz. Ama nasıl durursunuz? Forward'ın parlak makalesini okuyana kadar benim de aklıma gelmemişti: Görsel 6.8'in gösterdiği gibi, la-

zer yelkenin dış halkası ayrılır ve uzay aracının önüne hareket ederek lazer ışınını geriye yansıtmak suretiyle aracı ve küçük yelkenini yavaşlatır.[6] Yapılan hesaplar bu sistemin, normalde dört ışık yılı mesafedeki Centauri güneş sistemine insanları kırk yılda götürebileceğini gösteriyor. Oraya varıldığında da yeni bir dev lazer sistemi inşa edebilir ve Samanyolu galaksisindeki yıldızları gezmeyi sürdürebiliriz.

Peki, neden orada duralım ki? 1964'te Sovyet astronom Nikolay Kardaşev kullanıma soktukları enerjiye göre uygarlıkları notlamayı önermişti. Bir gezegenin, yıldızın (diyelim ki bir Dyson küresiyle birlikte) ve bir galaksinin enerjisini kullanmak Kardaşev ölçeklerine göre sırasıyla Tip I, Tip II ve Tip III uygarlıklara denk gelir. Sonraki düşünürler de tüm erişilebilir evreni kullanmayı Tip IV olarak nitelemişlerdir. O zamandan bu yana, hırslı yaşam formları için hem iyi hem kötü haberler vardır. Kötü haber, karanlık enerjinin varlığı ve gördüğümüz üzere, erişebileceğimiz alanı sınırlandırmasıdır. Carl Sagan gibi iyimser vizyonerler bile insanların başka galaksilere erişme ihtimallerini umutsuzlukla karşılamışlardır çünkü ışık hızına yakın seyahat etsek bile milyonlarca yıl sürecek bir yolculuğun ilk yüzyılında ölme eğilimindeyizdir. Pes etmeyi reddederek, yaşam sürelerini uzatmak için astronotları dondurmayı, yaşlanmalarını yavaşlatmak için ışık hızına çok yakın süratlerde seyahati ya da on binlerce nesil –insanlığın şu ana kadar var olduğundan daha uzun süre– boyunca seyahat edecek bir topluluğu göndermek gibi seçenekleri değerlendirdiler.

Süper zekâ bu resmi tamamıyla dönüştürerek galaksiler arası gezginler için her şeyi daha çok umut verici yapar. Hacimli insan yaşam destek sistemlerini taşıma ihtiyacını ortadan kaldırarak, YZ icadı teknolojiyi ekleyerek, galaksiler arası yerleşim çok daha kolay olur. Forward'ın lazer yelkenciliği uzay aracının tek bir

"tohum sondası" taşıyabilecek kadar büyük olmasının yeterli gelmesiyle çok daha ucuza gelir: Hedef güneş sisteminde bir asteroit ya da gezegene inerek sıfırdan yeni bir uygarlık kurma yetisine sahip bir robottur. Yanında direktifleri bile taşımasına gerek yoktur: İnşa etmesi gereken tek şey, ışık hızıyla ana uygarlığından detaylı çizimleri ve direktifleri alabilecek kadar büyük bir alıcı antendir. Bu da yapıldığında, yeni inşa edilmiş lazerlerini yeni tohum sondaları gönderip her seferinde bir yıldız sistemine yerleşerek galaksiyi kolonize etmeye devam etmek için kullanır. Galaksiler arasındaki geniş karanlık boşluk bile yol istasyonları olarak kullanılabilecek galaksiler arası bir dizi önemli yıldız (ana galaksilerinden çıkmış ıskartalar) içerme eğilimindedir. Bu da adadan adaya atlama stratejisini galaksiler arası lazer yelkenciliği için mümkün kılar.

Başka bir güneş sistemi ya da galaksi süper zeki YZ tarafından kolonileştirilirse, insanları oraya getirmek kolay olur. Elbette eğer insanlar YZ'ye bu amacı vermekte başarılı olurlarsa. İnsanlar hakkındaki tüm gerekli bilgi ışık hızında iletilir, sonrasında da YZ kuarkları ve elektronları istenilen insanlar olacak şekilde birleştirebilir. Bu, bir insanın DNA'sını belirlemek için gerekli iki gigabaytlık bilgiyi iletip, bebeği kuluçkaya alıp YZ tarafından yetiştirmek gibi düşük teknolojili bir yöntemle de yapılabilecekken, YZ'nin Dünya'daki orijinallerinden tüm hafızaları taranmış yetişkin insanları nano boyutta kuark ve elektronları birleştirerek meydana getirmesiyle de yapılabilir.

Bu da eğer bir zekâ patlaması olacaksa, ana sorunun galaksiler arası kolonileşmenin mümkün olup olmayacağı değil, ne kadar hızlı ilerleyeceği olduğu anlamına gelir. Yukarıda incelediğimiz tüm fikirler insanlardan geldiğine göre, yaşamın ne hızla genişleyebileceğine dair en alt sınırlar olarak değerlendirilmelidir. Hırslı bir süper zeki yaşam muhtemelen çok daha iyisini yapabilir

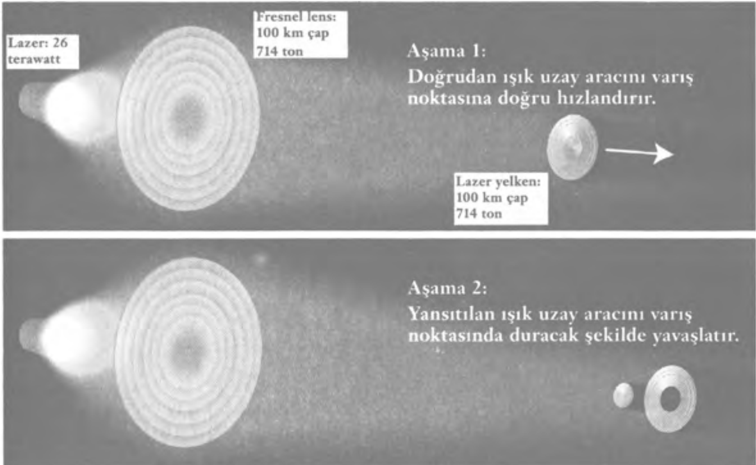
ve sınırları zorlamak için güçlü teşvikleri de olacaktır çünkü zaman ve karanlık enerjiye karşı olan yarışta ortalama yerleşme hızındaki her bir %1'lik artış %3 daha fazla galaksinin kolonize edilmesi anlamına gelecektir.

Mesela, lazer yelken sistemiyle on ışık yılı uzaktaki bir yıldız sistemine seyahat etmek yirmi yıl sürecekse ve sonrasındaki 10 yıl da oraya yerleşmek ve yeni lazerlerle tohum sondaları inşa etmekle geçecekse, uzayın kolonize edilmiş kesimi ortalama olarak ışık hızının üçte biri oranında tüm yönlere doğru büyüyen bir küre olacaktır. 2014 yılında kozmik olarak genişleyen uygarlıklar hakkında güzel ve derinlikli bir analizde, Amerikalı fizikçi Jay Olson ada atlama yaklaşımına daha yüksek teknoloji bir alternatifi ele almıştı. Bunda, iki farklı tür sonda vardı: *tohum sondalar* ve *genişleticiler*.[7] Tohum sondalar yavaşlayacak, konacak ve vardıkları yere yaşam tohumlarını ekeceklerdi. Genişleticiler ise, öte yandan, asla durmayacaklardı: Muhtelemen ramjet teknolojisinin gelişmiş bir versiyonunu kullanarak, uçarken madde toplayacak ve bu maddeyi hem yakıt hem de genişleticiler ve kendilerinin kopyalarını üretmek için kullanacaklardı. Bu kendi kendini kopyalayan genişletici filosu yakındaki galaksilere göre sabit bir hızı (mesela ışık hızının yarısını) korumak için yavaş yavaş hızlanmayı sürdürecek ve kabuk alanı başına sabit sayıda genişleticiyle küresel kabuğu genişletmeye yetecek sıklıkta kendilerini çoğaltacaklardı.

Son olarak da yukarıdaki her bir yöntemden daha hızlı genişlemeyi sağlayacak sinsi Kutsal Meryem yaklaşımı vardır: Bölüm 4'te bahsettiğimiz Hans Moravec'in "kozmik spam" yöntemi. Bir uygarlık, naif ve yeni evrim geçirmiş diğer uygarlıkları, kendilerini ele geçirecek bir süper zeki makine yapmaya kandıracak bir mesaj göndererek, özünde ışık hızında, yani baştan çıkarıcı mesajlarının kozmosta yayılma hızında genişleyebilir. İlerlemiş uygarlıkların gelecek ışık konileri içerisinde çoğu galaksiye ulaş-

ması için bu *tek* yol olabileceği için ve bunu denememek için sebepleri olmadığından dolayı, dünya dışından gelen her türlü iletime karşı oldukça şüpheli yaklaşmalıyız! Carl Sagan'ın kitabı *Mesaj*'da, biz Dünyalılar anlamadığımız bir makineyi inşa etmek için uzaylılardan gelen planları kullanıyorduk; bunu yapmayı önermiyorum...

Özet olarak, kozmik yerleşmeyi ele alan çoğu bilim insanı ve bilimkurgu yazarı, bana göre, süper zekânın imkânlarını görmezden gelmekte çok karamsar oluyorlar: İlgiyi yalnızca insan gezginlerle kısıtlayarak, galaksiler arası yolculuğun zorluğunu; ilgiyi yalnızca insanların icat ettiği teknolojiyle sınırlayarak da mümkün olanın fiziksel sınırlarına yaklaşmak için gereken süreyi abartıyorlar.



Görsel 6.8: Robert Forward'ın dört ışık yılı uzaklıktaki Centauri yıldız sistemine lazer yelken görevi için tasarımı. Başta Güneş sistemimizdeki güçlü lazer, uzay aracını lazer yelkenine radyasyon basıncı uygulayarak hızlandırır. Varış noktasına gelmeden önce frenlemek için de yelkenin dış kısmı ayrılır ve lazer ışığını uzay aracına geri yansıtır.

Kozmik Mühendislikle Bağlı Kalmak

Eğer karanlık enerji, son deneysel verinin gösterdiği üzere, uzak galaksileri birbirlerinden daha uzağa itmeyi sürdürürse, yaşamın geleceği için büyük bir sıkıntı olacaktır. Yani gelecekte bir uygarlık bir milyon galaksiye yerleşmeyi başarabilse bile, on milyarlarca yıl içinde karanlık enerji bu kozmik imparatorluğu birbiriyle iletişime geçemeyen binlerce farklı bölgeye ayıracaktır. Eğer gelecek yaşam bu bölünmeyi önlemek için hiçbir şey yapmazsa, yaşamın kalan en büyük kaleleri onları ayırmaya çalışan karanlık enerjinin üstesinden gelecek kadar güçlü toplam kütle çekimine sahip yaklaşık bin galaksi içerecek kümeler olacaktır.

Eğer süper zeki bir uygarlık farklı unsurlarıyla bağlantıda kalmak isterse, büyük ölçekli bir kozmik mühendislik yapması gerekecektir. Karanlık enerji onu tamamıyla erişimi dışına atmadan önce en büyük süperkümesine madde taşımak için ne kadar zamanı olacaktır? Büyük mesafeler boyunca bir yıldızı hareket ettirmenin bir yolu iki yıldızın dengeli biçimde birbirinin yörüngesinde olduğu ikili bir sisteme üçüncü bir yıldızı itmektir. Romantik ilişkilerde olduğu gibi üçüncü bir partnerin girişi işleri karıştırabilir ve üçünden birinin vahşice dışa atılmasına sebep olabilir; yıldızlar ölçeğinde bu, çok yüksek hızda gerçekleşir. Üç partnerden biri kara delikse, böyle dengesiz bir ilişki ev sahibi galaksinin dışına çıkaracak kadar hızla kütle fırlatmak için kullanılabilir. Maalesef, bu üç özneli teknik, yıldızlara, kara deliklere ya da galaksilere uygulanmış olsun, karanlık enerjinin üstesinden gelmek için gereken büyük uzaklıklara uygarlığın kütlesinin ancak ufacık bir kısmını hareket ettirebilir.

Ancak bu elbette ki süper zeki yaşamın, mesela dış galaksilerin kütlesinin çoğunu ana galaksi kümesine seyahat edebilecek uzay gemilerine dönüştürmek gibi, daha iyi yöntemlerle ortaya çıkmayacağı anlamına gelmez. Eğer bir sfalerizer inşa edilebilirse,

belki de maddeyi ışık olarak ana galaksi kümesine yollayabilecek enerjiye dönüştürmekte bile kullanılabilir, orada da maddeye geri döndürülebilir ya da güç kaynağı olarak kullanılabilir.

Birbirinden ne kadar ayrı olduklarından bağımsız olarak solucan deliğinin iki ucu arasında neredeyse anlık iletişim ve seyahati mümkün kılabilecek dengeli solucan delikleri inşa etmenin mümkün olduğu ortaya çıkarsa büyük şans olur. Bir solucan deliği uzayzamanda A noktasından B noktasına aradaki uzayı katetmeden gidebileceğiniz bir kestirmedir. Dengeli solucan delikleri Einstein'ın genel görelilik teorisine uygun olsa da ve *Mesaj* ile *Yıldızlararası* gibi filmlerde geçse de negatif yoğunluğa sahip garip varsayımsal bir tür maddenin varlığını gerektirir. Bu varlık da pek az anlaşılmış kuantum kütle çekimi etkilerine dayanır. Diğer bir deyişle, faydalı solucan delikleri imkânsız da olabilir ancak olmazlarsa, süper zeki yaşamın onları inşa etmek için ciddi bir sebebi olacaktır. Solucan delikleri sadece galaksiler arasında hızlı iletişimde devrim yapmakla kalmaz, dış galaksileri merkezdeki kümeye bağlayarak gelecek yaşamın idaresinin uzun mesafelerde bile bağlantılı kalabilmesini sağlar ve karanlık enerjinin iletişimi yok etmesine engel olur. İki galaksi stabil bir solucan deliğiyle bağlanınca, ne kadar uzağa giderlerse gitsin bağlı kalırlar.

Eğer, tüm kozmik mühendislik çabalarına rağmen, gelecekteki bir uygarlık bazı parçalarının sonsuza dek bağlantı dışına çıkmaya mahkûm olduğuna kanaat getirirse, gitmelerine izin verip şans dileyebilir. Ancak belli başlı zor sorulara cevap aramayı da içeren hırslı hesaplama amaçları varsa, yakıp yer açma stratejisine geçebilirler. Karanlık enerji tükenmiş kalıntılarını uzaklaştırıp götürmeden önce uzun süredir aradıkları cevapları ana galaksi kümesine yollayabileceği ümidiyle, çılgın bir hızla dış galaksilerin madde ve enerjilerini hesaplamaya dönüştüren devasa bilgisayarlara dönüştürebilirler. Bu yakıp yer açma stratejisi yalnızca "kozmik spam" metoduyla ulaşılabilecek çok uzak

bölgelerde uygulanabilecek, önceden orada yaşayanları üzecek olsa da uygun bir yöntemdir. Ana galaksi kümesinde, uygarlık mümkün olduğunca uzun sürecek maksimum koruma ve verimliliği hedefleyebilir.

Ne Kadar Uzun Süre Yaşayabilirsiniz?

Uzun ömür çoğu hırslı insanın, kurumun ve ulusun amaç edindiği bir şeydir. Eğer hırslı bir gelecek uygarlığı süper zekâyı geliştirir ve uzun ömür isterse, ne kadar uzun süre yaşayabilir?

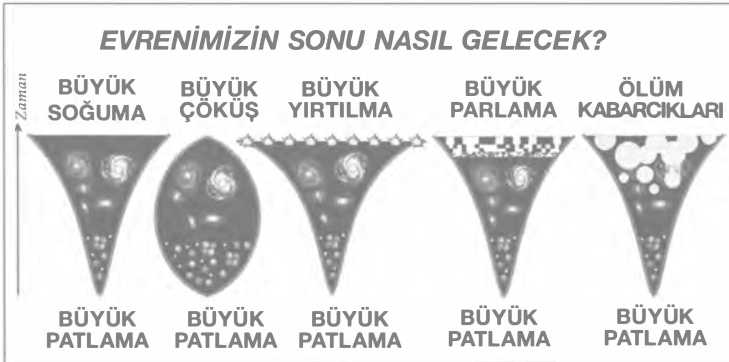
Uzak geleceğimizin ilk bütünlüklü bilimsel analizi elbette ki Freeman Dyson tarafından yapılmıştı ve Tablo 6.3 onun bazı ana bulgularını özetlemektedir. Sonuç olarak, zekâ araya girmezse, güneş sistemleri ve galaksiler zamanla yok olur, diğer her şey de onları takip eder. Geriye sonsuza dek sönecek bir radyasyon parıltısı bırakarak soğuk, ölü ve boş bir uzay ortaya koyar. Ancak Freeman analizini iyimser bir notla bitirir: "Yaşam ve zekânın evrenimize kendi amaçlarına göre şekil vermeyi başarması ihtimalini ciddiye almak için önemli oranda bilimsel sebebimiz vardır."[8]

Ne	Ne Zaman
Evrenimizin şimdiki yaşı	10^{10} yıl
Karanlık enerji galaksilerin çoğunu erişim dışına iter	10^{11} yıl
Son yıldızlar söner	10^{14} yıl
Gezegenler yıldızlardan ayrılır	10^{15} yıl
Yıldızlar galaksilerden ayrılır	10^{19} yıl
Yerçekimsel radyasyonla yörüngelerin çözülmesi	10^{20} yıl
Proton parçalanması (en erken)	$>10^{34}$ yıl
Yıldız kütleli kara delikler buharlaşır	10^{67} yıl
Süper kütleli kara delikler buharlaşır	10^{91} yıl
Tüm madde demire bozunur	10^{1500} yıl
Tüm madde kara deliklere dönüşür, sonra da buharlaşır	$10^{10^{26}}$ yıl

Tablo 6.3: Uzak gelecek tahminlerinin ikinci ve yedincisi dışında hepsi Freeman Dyson tarafından yapılmıştır. Bu hesaplamaları, 10^{10} ila 10^{11} yıl

içerisinde ortaya çıkabilecek birkaç çeşit "kozmokalips"i mümkün kılacak karanlık enerjinin keşfinden önce yapmıştı. Protonlar tamamen stabil kalabilirler; eğer olmazsa, deneyler yarısının bozunması için 10^{34} yıldan fazlasının gerektiğini söylüyor.

Bence süper zekâ Tablo 6.3'te listelenen sorunların çoğunu kolayca çözebilir çünkü maddeyi güneş sistemleri ve galaksilerden daha iyi bir şekilde yeniden düzenleyebilir. Mesela birkaç milyar yıl içinde Güneş'imizin ölümü gibi sık tartışılan zorluklar bizi engellemeyecek çünkü görece geri teknolojili bir uygarlık bile iki yüz milyar yıldan daha uzun ömürlü daha düşük kütleli yıldızlara kolayca taşınabilir. Süper zeki uygarlıkların yıldızlardan daha verimli enerji santralleri kuracağını varsayarsak, enerjiyi korumak için yıldız oluşumunu *önlemek* isteyebilirler: Bir yıldızın ana yaşam süreci boyunca tüm enerji çıktısını toplamak için bir Dyson küresi kullansalar bile (toplam enerjinin %0,1'ini elde ederek), iri yıldızların öldüğünde kalan %99,9'un israf edilmesini engelleyemeyebilirler. Ağır bir yıldız, enerjinin büyük kısmının sıyrılabilmeyi başaran nötrinolar olarak kaçtığı süpernova patlamasında ölür ve çok ağır yıldızlar için, büyük miktarda kütle kara deliği oluşturarak boşa gider, oradan da enerjinin sızması için 10^{67} yıl geçmesi gerekir.



Görsel 6.9: Evrenimizin on dört milyar yıl önce sıcak bir Büyük Patlama'yla başladığını, genişlediğini, soğuduğunu ve parçacıklarını atomları, yıldızları

ve galaksileri oluşturacak şekilde birleştirdiğini biliyoruz. Ama nihai kaderini bilmiyoruz. Önerilen senaryolar arasında bir Büyük Soğuma (sonsuz genişleme), bir Büyük Çöküş (içe çökme), bir Büyük Yırtılma (her şeyi yırtan hızda sonsuz genişleme), bir Büyük Parlama (çok genişlediğinde uzayın dokusunun ölümcül tanecikli doğasını açığa çıkarması) ve Ölüm Kabarcıkları (ışık hızıyla genişleyen ölümcül kabarcıklarda uzayın "donması") gibi seçenekler bulunur.

Madde/enerjisi bitmediği sürece süper zeki yaşam, yaşam alanını istediği durumda tutmayı başarabilir. Belki de bozunma sürecinin düzenli gözlemlerle yavaşlaması sayesinde, kuantum mekaniğinin *başında beklenen tencere etkisini* kullanarak protonların çözünmesini engellemenin bir yolunu bile bulabilir. Ancak gösteriyi sonlandıracak bir ihtimal de vardır: Tüm evreni on ila yüz milyar yıl gibi kısa bir süre içerisinde yok edecek bir "kozmokalips." Karanlık enerjinin keşfi ve sicim teorisindeki gelişme Freeman Dyson'ın ünlü makalesini yazarken bilmediği yeni kozmokalips senaryolarını ortaya çıkarmıştı.

Peki, şimdi evrenimiz, milyarlarca yıl sonra, nasıl sona erecek? Yaklaşan kozmik kıyametimiz ya da *kozmokalipsimiz* için beş ana şüphem var, bunları da Görsel 6.9'da belirttim: *Büyük Soğuma*, *Büyük Çöküş*, *Büyük Yırtılma*, *Büyük Parlama* ve *Ölüm Kabarcıkları.* Evrenimiz on dört milyar yıldır genişliyor. Büyük Soğuma, evrenimizin sonsuza kadar genişleyerek, kozmosumuzun soğuk, karanlık ve sonunda da ölü bir alana seyrelmesidir. Bu, Freeman makaleyi yazdığında en olası sonuç olarak görülüyordu. Bunu T. S. Eliot'ın seçeneği olarak düşünürüm: "İşte dünya böyle eriyor sona/Patlamayla değil, usulca." Eğer Robert Frost'un dünyanın sonunun buzla değil ateşle geleceği yaklaşımını seviyorsanız, kendinizi Büyük Çöküş'e hazırlayın. Bu senaryoda, kozmik genişleme tersine çevrilir ve her şey Büyük Patlama'nın tam tersine benzer bir şekilde sarsıntılı bir çöküş içinde birbirine

çarpmaya başlar. Son olarak Büyük Yırtılma da sabırsızlar için Büyük Soğuma'dır. Burada galaksilerimiz, gezegenlerimiz ve hatta atomlarımız bile sınırlı bir zaman sonrasındaki büyük bir finalde paramparça olurlar. Bu üçünden hangisine bahis oynarsınız? Bu da evrenimizin kütlesinin %70'ini oluşturan karanlık enerjinin uzay genişlemeyi sürdürürken ne yapacağına bağlıdır. Karanlık enerjinin, sırasıyla, değişmeden kalması, negatif yoğunluğa seyrelmesi ya da yüksek yoğunluğa yoğunlaşması durumlarına bağlı olarak Soğuma, Çöküş veya Yırtılma senaryolarından biri gerçekleşebilir. Karanlık enerjinin ne olduğuna dair hiçbir fikrimiz hâlâ olmadığına göre, size paramı nereye yatıracağımı söyleyeyim: %40'ını Büyük Soğuma'ya, %9'unu Büyük Çöküş'e, %1'ini de Büyük Yırtılma'ya yatırırdım.

Paramın kalan %50'si peki? Onu "yukarıdakilerin hiçbiri" seçeneğine saklıyorum çünkü biz insanların alçak gönüllü olması ve halen anlamadığımız bazı temel şeyler olduğunu kabul etmemiz gerektiğini düşünüyorum. Mesela, uzayın doğası. Soğuma, Çöküş ve Yırtılma sonlarının hepsi uzayın kendisinin stabil ve sonsuz oranda genişleyebilecek olduğunu varsayıyor. Uzayı kozmik dramanın ortaya çıktığı sıkıcı statik bir sahne olarak düşünürdük. Sonra Einstein bize uzayın ana aktörlerden biri olduğunu öğretti: Kara deliklere kıvrılabilir, yer çekimi dalgaları gibi dalgalanır ve genişleyen bir evren gibi esneyebilir. Hatta su gibi, bilinmez bir başka kozmokalips adayı olarak yeni fazın hızla genişleyen ölüm kabarcıklarıyla farklı bir faza doğru donabilir bile. Eğer Ölüm Kabarcıkları'nın var olmaları mümkünse, muhtemelen ışık hızıyla genişlerlerdi, tıpkı aşırı agresif bir uygarlıktan gelen artan kozmik spam küresi gibi.

Dahası, Einstein'ın teorisi uzayın esnemesinin devamlı sürerek evrenimizin Büyük Soğuma ve Büyük Yırtılma senaryolarında olduğu gibi sonsuz hacme varabilmesine izin verebileceğini

söylüyor. Bu doğru olamayacak kadar iyi gibi ve bence de pek mümkün değil. Paket lastiği iyi ve devamlı gibi gözükür, tıpkı uzay gibi ama çok fazla esnetirseniz kopar. Neden? Çünkü atomlardan yapılmıştır ve yeterince esnetmeyle bu taneli atomik lastik doğası önemli hale gelir. Uzayın da bizim fark edemeyeceğimiz kadar küçük bir ölçekte böyle bir taneliliğe sahip olması mümkün müdür? Kuantum yer çekimi araştırmaları 10^{-34}'ten daha küçük ölçeklerde bildiğimiz üç boyutlu uzaydan bahsetmenin mantıklı olmayacağını söylüyor. Eğer uzayın sarsıntılı bir "Büyük Parlama" sürecine girmeden sonsuz süre esneyemeyeceği gerçekten doğruysa, gelecek uygarlıklar uzayın ulaşabilecekleri en büyük genişlemeyen bölgesine (büyük bir galaksi kümesi) yerleşmeyi isteyebilirler.

Ne Kadar Çok Hesaplayabilirsiniz?

Gelecek yaşamın ne kadar *yaşayabileceğini* inceledikten sonra, şimdi de ne kadar yaşamak *isteyebileceğine* bakalım. Doğal olanın yaşayabileceğiniz kadar yaşamak istemek olduğunu düşünseniz bile, Freeman Dyson bu arzu için daha nicel bir sav sunmuştur: Hesaplama maliyeti yavaş hesapladığınızda düşer, bu yüzden işleri olabildiğince yavaştan alırsanız daha fazla iş halledebilirsiniz. Hatta Freeman evrenimizin sonsuza dek genişleyip soğumayı sürdürmesi durumunda sonsuz miktarda hesaplamanın mümkün olabileceğini hesaplamıştır.

Yavaş, sıkıcı anlamına gelmek zorunda değil: Eğer gelecek yaşam simüle edilmiş bir dünyada yaşarsa, öznel olarak deneyimlenmiş zaman akışının simülasyonun dış dünyada çalıştığı yavaşlıkla ilgisi olmak zorunda değil, bu yüzden de sonsuz hesaplama ihtimali simüle edilmiş yaşam formları için öznel ölümsüzlük anlamına gelebilir. Kozmolog Frank Tipler, sıcaklık ve yoğunluk

artarken hesaplamayı sonsuza doğru hızlandırarak Büyük Çöküş öncesindeki son anlarda nesnel ölümsüzlüğe ulaşılabileceği spekülasyonunu bu fikir üzerine kurmuştur.

Karanlık enerji hem Freeman'ın hem de Frank'in sonsuz hesap hayallerini suya düşürdüğü için, geleceğin süper zekâsı enerji tedarikini görece hızlı yakıp onları kozmik ufuk ve proton bozunması gibi problemlerle karşılaşmadan önce hesaplamaya dönüştürmeyi tercih edebilir. Eğer toplam hesaplamayı maksimize etmek nihai hedef ise, en iyi strateji çok yavaş (yukarıda bahsettiğimiz sorunlardan kaçınmak için) ile çok hızlı (hesaplama başına ihtiyaç duyulandan daha fazla enerji harcamak) arasında bir denge bulunması olacaktır.

Bu bölümde incelediğimiz her şeyi bir araya koyduğumuzda bu bize, en etkili enerji santralleri ve bilgisayarların süper zeki yaşama akıllara durgunluk verecek miktarda hesaplama yapma olanağı vereceğini söyler. On üç vattlık beyninize yüz yıl boyunca güç sağlamak yarım miligram madde kadar enerji gerektirir: tek bir şeker tanesinden daha az. Seth Lloyd'un çalışması, bir şeker tanesinin yaşamış tüm insanların ve binlerce kat fazlasının tüm yaşamlarının simülasyonuna güç sağlamasına olanak tanıyarak beynin bir katrilyon kat daha fazla enerji verimli yapılabileceğini ileri sürer. Eğer mevcut evrenimizdeki tüm madde insanları simüle etmek için kullanılabilirse bu, 10^{69} yaşamı mümkün kılardı ya da süper zeki YZ hesaplama gücüyle ne yapmayı tercih ederse onu. Eğer simülasyonlar daha yavaş çalıştırılırsa daha fazla yaşam bile mümkün olabilir.[9] Öte yandan, kitabı *Superintelligence*'ta Nick Bostrom enerji verimliliği hakkında daha tasarruflu varsayımlarla 10^{58} insan yaşamının simüle edilebileceğini tahmin ediyor. Ancak bu rakamları incelediğimizde, yaşamın gelecekte gelişme potansiyelini çarçur etmeyeceğimizden emin olma sorumluluğumuz kadar yüksek olduğunu görürüz. Bostrom'un da dediği

gibi: "Eğer tüm bir yaşamda deneyimlenmiş tüm mutluluğu bir sevinç gözyaşıyla temsil edersek, bu ruhların mutluluğu Dünya'nın okyanuslarını her saniye doldurup tekrar doldurabilir ve bunu yüz milyar milyar bin yıllar boyunca yapmayı sürdürebilir. Bizim yapmamız gereken şey ise bunların tamamen sevinç gözyaşı olmasını sağlamaktır."

Kozmik Hiyerarşiler

Işık hızı yalnızca yaşamın yayılmasını değil iletişime, bilince ve kontrole güçlü sınırlar koyarak ayrıca yaşamın doğasını da sınırlar. O halde, kozmosumuzun büyük bir kesimi canlı hale gelirse yaşam neye benzeyecektir?

Düşünce Hiyerarşileri

Bir sineği elinizle ezmeye çalışıp başaramadığınız oldu mu? Sizden daha hızlı tepki gösterebilmesinin sebebi daha küçük olmasıdır, bu yüzden de gözleri, beyni ve kasları arasında bilginin seyahat etmesi daha az zaman alır. Bu "daha büyük = daha yavaş" ilkesi, sadece hız sınırının elektrik sinyallerinin nöronlar arasındaki hareket hızıyla belirlendiği biyolojide değil, eğer hiçbir bilgi ışıktan hızlı hareket edemiyorsa gelecek kozmik hayatta da geçerlidir. Bu yüzden de zeki bilgi işlem sistemleri için büyümek ilginç bir değiş tokuşu da beraberinde getiren karmaşık bir fırsattır. Bir tarafta, büyümek daha daha fazla partiküle sahip olmasına sebep olur ve daha karmaşık düşünceye izin verir. Öte yandan, ilgili bilginin tüm kısımlara yayılmasını geciktirdiği için, tamamen küresel düşüncelere sahip olma hızını azaltır.

Yani eğer yaşam tüm kozmosa dağılırsa, hangi formu tercih eder: Basit ve hızlı olanı mı yoksa karmaşık ve yavaş olanı mı? Ben Dünya yaşamının yaptığının aynısını yapacağını düşünüyorum:

ikisini de! Dünyanın biyosferi iki yüz tonluk devasa mavi balinadan minnacık ama dünyanın tüm balıklarının toplamından daha fazla biyokütleye sahip olan 10^{-16} kilogramlık bakteri *Pelagibacter*'e kadar inanılmaz bir çeşitliliğe sahiptir. Dahası büyük, karmaşık ve yavaş olan organizmalar basit ve hızlı küçük modüller içererek tembelliklerini sık sık azaltırlar. Mesela, göz kırpma refleksiniz beyninizin büyük kısmını kapsamayan küçük ve basit bir devreyle çalıştırıldığı için oldukça hızlıdır: Eğer ezmesi zor sinek kaza eseri gözünüze doğru ilerlerse, saniyenin onda birinde, ilgili bilginin beyninize yayılıp ne olduğuna dair bilinçli bir çıkarsama yapma fırsatı bulamadan göz kırparsınız. Bu bilgi işlemini hiyerarşi modülüne organize ederek, biyosferimiz hem karnını doyurup hem de pastasının durmasını, hem hıza hem de karmaşıklığa ulaşmayı başarır. Biz insanlar çoktan bu aynı hiyerarşik stratejiyi paralel hesaplamayı optimize etmek için kullanıyoruz.

Dâhilî iletişim yavaş ve maliyetli olduğu için, gelecekteki ilerlemiş kozmik yaşamın da aynı şeyi yapmasını beklerim; böylece hesaplamalar mümkün olduğunca yerelde yapılacaktır. Eğer bir hesaplama bir kiloluk bilgisayarla yapılacak kadar basitse, galaksi boyutunda bir bilgisayara yaptırılması ters etkili olur çünkü her hesaplama adımından sonra paylaşılacak bilginin ışık hızında gelmesini beklemek her adım için yüz bin yıl gibi komik bir gecikmeye sebep olur.

Gelecekteki bu bilgi işleminin bir bölümünün subjektif bir deneyim yaşayarak bilinçli olup olmayacağı tartışmalı ve ilginç bir konudur ve bunu Bölüm 8'de inceleyeceğiz. Eğer bilinç, sistemin farklı kısımlarının bir diğeriyle iletişim kurma yetisine ihtiyaç duyarsa, büyük sistemlerin düşünceleri zorunlu olarak daha yavaş olur. Siz ya da geleceğin Dünya büyüklüğündeki süper bilgisayarı saniye başına mümkün olduğunca çok düşünceye sahip olurken, galaksi boyutundaki bir zihin ancak yüz bin yılda bir bir düşün-

ceye sahip olabilir ve milyar ışık yılı boyutundaki kozmik bir zihin karanlık enerji onu bağlantısız parçalara ayırmadan önce toplamda yaklaşık on tane düşünceye sahip olabilir. Öte yandan, bu azıcık ama kıymetli düşünceler ve beraberindeki deneyimler oldukça derin olabilir!

Kontrol Hiyerarşileri

Eğer düşünce geniş çapta bir hiyerarşide organize edildiyse, peki ya o zaman güç? Bölüm 4'te, zeki varlıkların doğal olarak kendilerini Nash denklemi güç hiyerarşilerinde organize ettiklerini ve stratejisini değiştiren herhangi bir varlığın durumunun daha kötüye gittiğini görmüştük. İletişim ve ulaştırma teknolojisi iyileştikçe, bu hiyerarşiler de büyür. Eğer süper zekâ bir gün kozmik ölçeğe genişlerse, güç hiyerarşileri neye benzer? Serbest ve merkezî olmayan bir şekilde mi yoksa tamamen otoriter mi olur? İşbirliği yoğunlukla karşılıklı faydaya mı yoksa baskı ve tehditlere mi dayanır?

Bu soruları aydınlatmak için, şimdi hem havuç hem de çubuğu inceleyelim: Kozmik ölçüde işbirliği için ne gibi teşvikler var ve ne gibi tehditler bunun uygulanmasını sağlayabilir?

Havuçla Kontrol

Ticaret, işbirliğinin geleneksel lokomotifi olmuştur çünkü şeyleri üretmenin görece zorluğu tüm gezegende farklılık gösterir. Eğer bir kilogram gümüş çıkarmak bir bölgede bir kilogram bakır çıkarmaktan üç yüz kat daha maliyetli ama başka bir bölgede yüz kat daha maliyetliyse, iki yüz kilogram bakırı bir kilogram gümüşle takas ederlerse ikisi de kârlı çıkacaktır. Eğer bir bölgenin çok daha yüksek teknolojisi varsa, ikisi de yüksek teknoloji malları hammaddeyle değiş tokuş etmekten fayda sağlayabilir.

Ancak, süper zekâ temel parçacıkları istenen herhangi bir formda yeniden düzenleyebilecek teknolojiyi geliştirirse, uzun mesafe ticareti için gerekli itkinin çoğu ortadan kalkar. Parçacıklarını yeniden düzenleyerek bakırı gümüşe dönüştürmek daha kolay ve hızlıyken, uzak güneş sistemleri arasında gümüş ticareti yapmakla neden uğraşasın ki? Hem bilgisi hem de hammaddesi (herhangi bir materyal olabilir) iki yerde de mevcutken neden yüksek teknoloji makineleri galaksiler arasında gezdirelim ki? Bence süper zekâyla dolup taşan bir kozmosta, uzun mesafe taşınmaya değer tek meta *bilgi* olacaktır. Bunun tek istisnası kozmik mühendislik projeleri için kullanılacak madde olabilir; örneğin yukarıda bahsedilen, karanlık enerjinin uygarlıkları birbirinden ayırma eğilimine karşı gelmek için. Geleneksel insan ticaretinin aksine, bu madde herhangi bir biçimde, hatta bir enerji ışını olarak bile sevk edilebilir çünkü bu maddeyi alacak olan süper zekâ onu istediği türde bir madde olarak yeniden kurabilir.

Eğer bilginin paylaşılması ve ticaretinin yapılması kozmik işbirliğinin ana itici gücü olacaksa, bunlar ne tür bilgiler olacaktır? Üretmesi devasa ve zaman harcayan hesaplamalar gerektiren her türlü bilgi değerli olacaktır. Bir süper zekâ fiziksel gerçekliğin doğası, teoremler ve optimal algoritmalar hakkında zor matematik sorularına ve mükemmel teknolojiyi en iyi şekilde inşa etmenin yollarıyla ilgili zor mühendislik sorularına cevaplar isteyebilir. Hazcı yaşam formları muazzam dijital eğlence ve simüle edilmiş deneyimler isteyebilir ve kozmik ticaret bitcoin benzeri kozmik bir kripto para için talep yaratabilir.

Böyle paylaşım fırsatları yalnızca yaklaşık eş güçlere sahip varlıklar arasında değil, Güneş sistemi büyüklüğünde nodlar ile galaksiler ağı ya da galaksi büyüklüğünde nodlar ile kozmik bir ağ arasındaki güç hiyerarşileri arasında da bilgi akışını teşvik edebilir. Bu nodlar bunu daha büyük bir şeyin parçası olmanın

mutluluğu, kendi başlarına geliştiremeyecekleri teknolojiler ve cevapların sağlanması ve dış tehditlere karşı savunma için isteyebilirler. Yedekleme yoluyla neredeyse ölümsüzlük vaadini de değerlendiriyor olabilirler: Tıpkı çoğu insanın fiziksel bedenleri öldükten sonra zihinlerinin yaşamayı sürdüreceği inancında teselli bulması gibi, gelişmiş bir YZ de orijinal fiziksel donanımı enerji rezervlerini bitirdikten sonra bir süper bilgisayar ağında zihninin ve bilgisinin yaşamasına kıymet verebilir.

Öte yandan ağ da nodlardan sonuçların acilen gerekmediği uzun vadeli devasa hesaplama görevlerine yardımcı olmalarını isteyebilir. Böylece cevaplar için binlerce ya da milyonlarca yıl beklemeye değer. Yukarıda konuştuğumuz gibi, ağ da nodlarının, galaksi boyutunca kütle yoğunlaşmalarını beraber hareket ettirerek yıkıcı karanlık enerjiye karşı gelmek gibi devasa kozmik mühendislik projelerini yerine getirmesini isteyebilir. Eğer seyahat edilebilir solucan deliklerinin mümkün ve inşa edilebilir olduğu ortaya çıkarsa, o zaman bir ağın ilk önceliği imparatorluğunu sonsuza dek bağlı tutabilmek ve karanlık enerjiye engel olmak için onlardan bir ağ kurmak olacaktır. Kozmik bir süper zekânın ne gibi nihai hedefleri olacağı sorusu da Bölüm 7'de inceleyeceğimiz baş döndürücü ve ihtilaflı bir konudur.

Sopayla Kontrol

Dünya imparatorlukları genelde tebaalarını havuç ve sopa yöntemlerini kullanarak işbirliğine mecbur kılar. Roma İmparatorluğu'nun tebaası işbirliği karşılığında ödül olarak onlara sunulan teknoloji, altyapı ve savunmaya değer veriyorlardı. Ayrıca vergileri ödememenin ya da isyan etmenin yaratacağı kaçınılmaz tepkiden de korkuyorlardı. Roma'dan dış eyaletlere asker göndermek için uzun zaman gerektiği için, göz korkutma kısmen neredeyse anlık cezalar

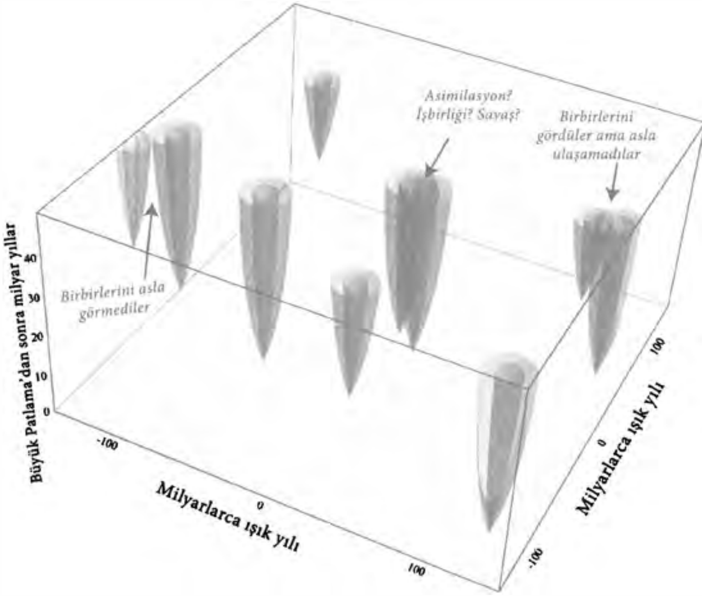
vermek için güçlendirilmiş yerel kuvvetlere ve sadık görevlilere aktarılmıştı. Süper zeki bir ağ, kozmik imparatorluğu boyunca yarattığı sadık muhafızlar ağını kullanarak benzer bir stratejiyi uygulayabilir. Süper zeki tebaaları kontrol etmesi zor olduğu için, en kolay uygulanabilir strateji görece aptal olarak %100 sadık olacak biçimde programlanmış, yalnızca kuralların uyulup uyulmadığını izleyerek uyulmadığı takdirde otomatik olarak kıyamet cihazını tetikleyecek YZ muhafızlarını kullanmak olabilir.

Diyelim ki ağ YZ, beyaz bir cücenin kontrol etmek istediği güneş sistemi boyutundaki uygarlığın çevresine yerleştirilmesini ayarlasın. Beyaz cüce görece büyük bir yıldızın tükenmiş kabuğudur. Büyük miktarda karbon içeren bu cisim gökte devasa bir elması andırır ve Dünya'dan daha küçükken Güneş'ten daha ağır olabilecek kadar yoğundur. Hindistanlı fizikçi Subrahmanyan Chandrasekhar'ın da bilinen bir şekilde kanıtladığı gibi eğer *Chandrasekhar sınırını*, yani Güneş'imizin kütlesinin 1,4 katını, geçene dek cisme kütle eklemeyi sürdürürseniz, tip A1 süpernova olarak bilinen sarsıntılı bir termonükleer patlama ortaya çıkar. Eğer YZ ağı, bu beyaz cüceyi Chandrasekhar sınırına çok yakın olacak şekilde umursamaksızın ayarlarsa, muhafız YZ ileri derecede aptal olsa bile (gerçekten de çok aptal olduğu için) etkili olabilir: Yalnızca hükmedilen uygarlığın aylık kozmik bitcoin, matematik kanıt ya da şart koşulan diğer vergi kotalarını iletip iletmediğini doğrulamak ve iletilmediyse, beyaz cüceye süpernovaya sebep olacak kadar kütle göndererek tüm bölgeyi yok edecek biçimde programlanabilir.

Galaksi boyutunda uygarlıklar, yüksek sayıda yoğun nesneleri galaksi merkezindeki canavar kara deliğe yakın bir yörüngeye yerleştirerek ve bu kütleleri, mesela çarpıştırarak, gaza dönüştürmekle tehdit ederek benzer biçimde de kontrol edilebilir. Bu gaz sonrasında kara deliği beslemeye başlayarak onu güçlü bir

kuasara dönüştürür ve galaksinin büyük bir kısmını potansiyel olarak yaşanamaz kılar.

Özet olarak, gelecek yaşamın kozmik uzaklıklar üzerinden işbirliği yapması için güçlü motivasyonları vardır ancak böyle bir işbirliğinin temelde karşılıklı faydaya mı yoksa korkunç tehditlere mi dayanacağı ucu açık bir sorudur. Fiziğin empoze ettiği sınırlar iki senaryoya da izin vermektedir, bu yüzden de sonuç üstün gelen hedeflere ve değerlere bağlı olacaktır. Gelecek yaşamın bu hedefler ve değerlerini etkileme yetimizi Bölüm 7'de inceleyeceğiz.



Görsel 6.10: Eğer yaşam uzayzamanda (yer ve zaman olarak) çoklu noktalarda bağımsız olarak evrimleşir ve uzayı kolonize etmeye başlarsa, uzay her biri Görsel 6.7'deki şampanya bardağına benzeyen genişleyen kozmik biyosferler ağı içerir. Her biyosferin tabanı, kolonileşme başladığındaki yeri ve zamanı temsil etmektedir. Mat ve şeffaf şampanya bardakları sırasıyla ışık hızının %50 ve %100'ünde kolonileşmeye denk düşer ve örtüşmeler bağımsız uygarlıkların buluştuğu yerleri gösterir.

Uygarlıklar Çarpıştığında

Şu ana dek, yaşamın bizim kozmosumuza tek bir zekâ patlamasından gelerek yayıldığı senaryoları tartıştık. Peki, yaşam birden fazla yerde bağımsız bir biçimde evrimleşirse ve iki uygarlık karşılaşırsa ne olur?

Rastgele bir güneş sistemini düşünürsek, yaşamın gezegenlerinden birinde evrimleşmesi, gelişmiş teknoloji geliştirmesi ve uzaya yayılması ihtimali mevcuttur. Bu olasılık teknolojik yaşam Güneş sistemimizde çoktan geliştiği ve fizik kanunları uzaya yerleşmeye izin verdiği için sıfırdan büyük görünür. Eğer uzay yeterince büyükse (zaten, kozmolojik enflasyon teorisi engin ya da sonsuz olduğunu söyler), Görsel 6.10'da gösterildiği üzere, böyle çok fazla genişleyen uygarlık vardır. Jay Olson'ın yukarıda bahsettiğimiz makalesi böyle genişleyen kozmik biyosferlerin incelikli bir analizini de içerir ve Toby Ord Future of Humanity Institute'ta meslektaşlarıyla benzer bir analiz gerçekleştirmiştir. Üç boyutta incelendiğinde, bu kozmik biyosferlerdeki uygarlıklar tüm yönde aynı hızda genişlediği müddetçe kelimenin tam anlamıyla kürelerdir. Uzayzamanda, Görsel 6.7'deki şampanya bardağının üst kısmı gibi görünürler çünkü karanlık enerji eninde sonunda her uygarlığın ulaşabileceği galaksi sayısını sınırlar.

Eğer uzaya yerleşen komşu uygarlıklar arasındaki uzaklık karanlık enerjinin genişlemelerine izin verdiğinden çok daha büyükse, birbirleriyle asla iletişime geçemezler hatta varlıklarından bile haberdar olamazlar, bu yüzden de kozmosta yalnız olduklarını hissedeceklerdir. Eğer kozmosumuz daha bereketliyse ve komşular birbirine daha yakınsa, bazı uygarlıklar illaki çakışacaktır. Bu çakışan bölgelerde ne olur? İşbirliği mi, rekabet mi yoksa savaş mı?

Avrupalılar Afrika ve Amerika'yı fethedebilmişlerdi çünkü üstün teknolojileri vardı. Tersine, iki süper zeki uygarlığın bir-

biriyle karşılaşmasından çok daha önce, teknolojilerinin aynı seviyeye gelmesi, yalnızca fizik kurallarıyla sınırlanmış olması akla yatkındır. Bu da bir süper zekânın diğerini istese bile kolayca fethedemeyeceği anlamına gelir. Dahası, eğer amaçları görece aynı gibi olursa, fetih ya da savaşı istemek için pek sebepleri de olmaz. Örneğin, eğer ikisi de mümkün olduğunca çok güzel teorileri kanıtlamaya ve zekice algoritmaları icat etmeye çalışıyorlarsa, bulgularını paylaşabilir ve ikisi de daha büyük kazanç sağlar. Sonuçta, bilgi insanların kavga ettikleri kaynaklardan oldukça farklıdır. Çünkü bilgiyi aynı anda hem verebilir hem de tutabilirsiniz.

Bazı yayılmacı uygarlıklar temelde değişmez hedeflere sahiptirler: tıpkı bir köktenci tarikat ya da salgın bir virüs gibi. Ancak bazı gelişmiş uygarlıkların açık görüşlü insanlara daha çok benziyor olması da akla yatkındır: Yeterince ikna edici savlarla karşılaştıklarında hedeflerini düzenlemeye razıdırlar. Eğer ikisi buluşursa, silahların değil fikirlerin çatışması olur, en ikna edici olan başarıya ulaşır ve diğer uygarlık tarafından kontrol edilen bölgede amaçlarını ışık hızıyla yayar. Komşularınızı asimile etmek kolonileşmekten daha hızlı bir genişleme stratejisidir çünkü etki alanınız fikirlerin hareket ettiği hızda yayılır (telekomünikasyonu kullanan ışık hızıyla). Fiziksel kolonileşme ise kaçınılmaz olarak ışık hızından yavaş olacaktır. Bu asimilasyon *Star Trek*'teki Borg'un kullandığı şekilde zorla değil, fikirlerin ikna edici üstünlüğü ve asimile olanın daha iyi durumda olmasıyla gönüllü biçimde olur.

Gelecek kozmosun hızla genişleyen iki tür kabarcığı içerebileceğini gördük: genişleyen uygarlıklar ile ışık hızında genişleyen ve tüm temel parçacıklarımızı yok edip uzayı yaşanmaz kılan Ölüm Kabarcıkları. Hırslı bir uygarlık da bu sebeple üç tip bölgeyle karşılaşabilir: ıssız bölgeler, yaşam kabarcıkları ve ölüm kabarcıkları. Eğer işbirliğine kapalı rakip uygarlıklardan

korkarsa, hızlı bir "toprak kapma" yarışı başlatmak ve ıssız alanları rakiplerinden önce kolonileştirmek için güçlü motivasyonları olur. Ancak, başka uygarlıklar olmasa bile, yalnızca karanlık enerji onları erişilemez kılmadan önce kaynakları ele geçirmek için aynı yayılmacı motivasyona sahip olabilir. Başka bir yayılmacı uygarlığa rastlamanın, komşumuzun ne kadar işbirlikçi ve açık görüşlü olmasına bağlı olarak boş bir uzay alanına denk gelmekten nasıl daha iyi ya da daha kötü olabileceğini gördük. Ancak Ölüm Kabarcığı yerine herhangi bir yayılmacı uygarlığa (uygarlığınızı ataşlara dönüştürmeye çalışıyor olsalar bile) rastlamak daha iyidir. Çünkü Ölüm Kabarcığı onunla savaşsanız da anlaşmaya çalışsanız da ışık hızıyla genişlemeye devam eder. Bizim Ölüm Kabarcığı'na karşı tek korumamız uzaktakilerin bize ulaşmasından bizi koruyan karanlık enerjidir. Bu yüzden eğer Ölüm Kabarcıkları sık görülüyorsa, karanlık enerji bizim düşmanımız değil dostumuzdur.

Yalnız mıyız?

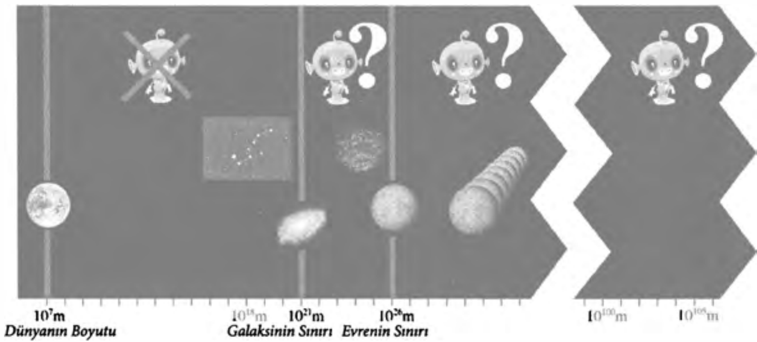
Pek çok insan evrenimizin büyük bir kısmında gelişmiş yaşam olduğundan emin, bu yüzden de kozmik perspektifte insanlığın soyunun tükenmesi çok da önemli değil. Sonuçta ilham verici *Star Trek* tarzı bir uygarlık birden gelip Güneş sistemimize yeniden yaşam getirir, ileri teknolojileriyle bizi yeniden kurup canlandıracaklarsa, silinip gideceğiz diye neden endişe edelim? Ben bu *Star Trek* varsayımını tehlikeli buluyorum çünkü bizi yanlış bir güvenlik hissiyatına kaptırıp uygarlığımızı duyarsız ve umursamaz yapar. Gerçekten de evrenimizde yalnız olmadığımız varsayımının yalnızca tehlikeli değil, ayrıca muhtemelen de yanlış olduğunu düşünüyorum.

Bu az insanın paylaştığı bir görüş* ve gayet yanlış düşünüyor olabilirim ancak kesinlikle savuşturamayacağımız bir olasılık ve bize dikkat etmek ve uygarlığımızın sonunu getirmemek için ahlaki bir sorumluluk yüklüyor.

Kozmoloji üzerine dersler verdiğimde, sık sık dinleyicilerden evrenimizin (Büyük Patlama'dan bu yana geçen 13.8 milyar yılda ışığın bize ulaştığı uzay kesimi) başka bir yerinde zeki yaşam olduğunu düşünenlerin el kaldırmasını isterim. İstisnasız biçimde neredeyse herkes el kaldırır, anaokulu çocuklarından kolej öğrencilerine kadar. Neden diye sorduğumda, aldığım cevap evrenimizin çok büyük olduğu ve en azından istatistiksel olarak bir yerinde yaşamın olması gerektiğidir. Bu sava yakından bakalım ve zayıflığını saptayalım.

Her şey tek bir sayıya indirgenir: Görsel 6.10'daki bir uygarlık ve en yakın komşusu arasındaki tipik uzaklık. Eğer bu uzaklık 20 milyar ışık yılından daha fazlaysa, evrenimizde (Büyük Patlama'dan bu yana geçen 13.8 milyar yılda ışığın bize ulaştığı uzay kesimi) yalnız olmayı ve uzaylılarla asla karşılaşmamayı bekleyebiliriz. Peki, bu uzaklıktan ne beklemeliyiz? Hiçbir fikrimiz yok. Bu, komşularımıza uzaklığımızın yaklaşık 1000 … 000 metre olduğunu ve toplam sıfır sayısının gayet 21, 22, 23, …, 100, 101, 102 ya da daha fazla olduğu anlamına gelir ama muhtemelen 21'den pek de az değildir çünkü uzaylılara dair ikna edici bir kanıt göremedik (bkz. Görsel 6.11). En yakın komşu uygarlığın, yarıçapı 10^{26} metre olan evrenimizde olması için, sıfırların sayısının 26'yı geçmemesi gerekir ve 22 ile 26 arasındaki dar alana sıfırların sayısının düşme ihtimali de görece küçüktür. Bu yüzden de evrenimizde yalnız olduğumuzu düşünüyorum.

* Ancak John Gribbin 2011'de yayımladığı kitabı *Alone in the Universe*'te benzeri bir sonuca varır. Bu soru üzerine ilgi çekici bir dizi bakış açısı için, Paul Davies'in 2011 tarihli kitabı *The Eerie Silence*'ı da tavsiye ederim.



Görsel 6.11: Yalnız mıyız? Yaşam ve zekânın nasıl oluştuğuna dair büyük belirsizlikler, uzaydaki en yakın komşu medeniyetin yukarıdaki yatay eksende olabileceğini göstermektedir; bu da onun galaksimizin sınırı (yaklaşık 10^{21} metre) ile evrenimizin sınırı (yaklaşık 10^{26} metre) arasındaki dar menzilde olmasının düşük bir ihtimal olduğunu göstermektedir. Bu menzilden çok daha yakında olsaydı, galaksimizde o kadar çok gelişmiş medeniyet olurdu ki şimdiye dek onları fark ederdik. Bu da evrende yalnız olduğumuz ihtimalini güçlendirmektedir.

Bu savın detaylı bir ispatını *Our Mathematical Universe* isimli kitabımda yaptım, bu yüzden de burada yeniden ele almak istemiyorum ancak komşu uzaklığı hakkında fikrimizin olmamasının temel sebebi belirli bir noktada zeki yaşamın ortaya çıkması ihtimali hakkında da fikrimizin olmamasıdır. Amerikalı astronom Frank Drake'in de işaret ettiği gibi, bu olasılık orada yerleşilebilir bir ortam olma olasılığı (uygun bir gezegeni), orada yaşamın oluşma olasılığı ve bu yaşamın zeki olma olasılığının çarpılmasıyla hesaplanır. Lisansüstü öğrencisiyken, bu üç olasılığın herhangi birisi hakkında bilgimiz yoktu. Son yirmi yılda, başka yıldızların yörüngesinde dönen gezegenlerin keşifleriyle birlikte, yalnızca bizim galaksimizde milyarlarca olan gezegenler arasında çok fazla yerleşilebilir gezegen olması mantıklı gözükmektedir. Yaşamın ve sonrasında da zekânın gelişmesi ihtimali ise halen belirsizdir: Bazı uzmanlar birinin ya da ikisinin de kaçınılmaz olduğunu ve en

yaşanabilir gezegenlerde gerçekleştiğini düşünüyorlar. Diğerleri ise birinin ya da ikisinin de aşırı derecede nadir olduğunu çünkü evrimsel darboğazların geçilmesi için çok büyük şans gerektiğini ileri sürüyorlar. Bazıları bu darboğazların kendi kendini kopyalayan yaşamın erken aşamalarında tavuk-yumurta problemi içerdiklerini söylerler: Mesela, modern bir hücrenin, kodumuzu okuyup proteinlerimizi inşa eden hayli karmaşık moleküler makine olan bir ribozom inşa etmesi için, başka bir ribozoma ihtiyacı vardır ve ilk ribozomun kendinden daha basit bir şeyden yavaş yavaş evrimleştiği de kesin değildir.[10] Diğer darboğazlar ise yüksek zekânın gelişmesini içerir. Örneğin, dinozorlar Dünya'ya biz insanlardan bin kat daha fazla, yüz milyon yıldan uzun süre hüküm sürmüş olsalar da evrim kaçınılmaz biçimde onları yüksek zekâya sürükleyip teleskop ve bilgisayarları icat ettirmedi.

Bazı insanlar benim argümanıma, evet, zeki yaşam nadir *olabilir* ama aslında değildir diyerek karşı çıkıyorlar. Galaksimiz ana akım bilim insanlarının fark etmediği zeki yaşamla dolup taşmaktadır. UFO meraklılarının iddia ettiği gibi, belki de uzaylılar çoktan Dünya'yı ziyaret ettiler. Belki de uzaylılar Dünya'yı ziyaret etmediler ama orada bir yerlerdeler ve bizden bilinçli olarak saklanıyorlar (bu ABD'li astronom John A. Ball tarafından "hayvanat bahçesi hipotezi" diye adlandırılır ve Olaf Stapledon'ın *Star Maker* kitabı gibi bilimkurgu klasiklerinde de geçer). Belki de bilinçli olarak saklanmadan orada bir yerlerde duruyorlar: uzaya yerleşmekle ya da bizim fark edeceğimiz büyük mühendislik projeleriyle ilgilenmiyorlar.

Elbette bu olasılıklara karşı açık görüşlü olmak zorundayız ancak bunların hiçbirine dair genel olarak kabul edilmiş bir kanıt olmadığı için, alternatifi de ciddiye almamız gerekir: yalnızız. Dahası, hepsinin fark edilmelerini sağlayan hedefleri olduğunu varsayarak ilerlemiş uygarlıkların çoğulluğunu hafife almamamız gerektiğini düşünüyorum: Yukarıda kaynak toplamanın bir

uygarlığın sahip olması için doğal bir hedef olduğunu görmüştük ve bizim fark etmemiz için gereken tek şey ise *bir* uygarlığın açıktan açığa yerleşebileceği her yere yerleşmeye karar vermesi ve galaksimiz ile ötesini yutuvermesidir. Galaksimizde Dünya'dan milyarlarca yıl yaşlı milyonlarca yaşanabilir Dünya benzeri gezegen olduğu gerçeğiyle yüzleşince, galaksiye yerleşmeleri için hırslı sakinlere yeterli zaman verirsek, şu bariz yorumlamayı görmezden gelmemiz imkânsız olur: Hayatın başlangıcı öylesine rastgele bir talih gerektirir ki yaşanabilir gezegenlerin iskân edilmiş olması çok düşük ihtimaldir.

Eğer yaşam nadir *değilse*, yakında öğreneceğiz. Hırslı astronomik araştırmalar Dünya benzeri gezegenlerin atmosferlerini yaşam tarafından üretilen oksijenin kanıtları için araştırıyor. *Herhangi* bir yaşam için yapılan bu araştırmaya paralel olarak, *zeki* yaşamın araştırılması da yakın zaman önce Rus hayırsever Yuri Milner'in yüz milyon dolarlık "Breakthrough Listen" projesiyle canlandırıldı.

Gelişmiş yaşam ararken ileri seviye insan merkezli olmamak önemlidir: Eğer bir dünya dışı uygarlık keşfedersek, süper zeki olma ihtimali yüksektir. Martin Rees'in yeni bir makalesinde söylediği gibi, "İnsan teknolojik uygarlığının tarihi yüzyıllarla ölçülür ve insanların inorganik zekâ tarafından ele geçirilmesine ya da aşılmasına yalnızca bir ya da iki yüzyıl daha kalmış olabilir. Sonrasında da bu zekâ, milyarlarca yıl boyunca sürecek evrimine devam edecektir... organik formundaki o kısa anda onu 'yakalamamız' pek olası değil."[11] Jay Olson'ın yukarıda bahsettiğimiz uzaya yerleşme makalesinde yaptığı çıkarıma katılıyorum: "Gelişmiş zekânın, evrenin kaynaklarını yalnızca mevcut dünya benzeri gezegenleri insanların gelişmiş versiyonlarıyla doldurmak için kullanacağı ihtimalini teknolojinin ilerlemesinin olası olmayan bir bitiş noktası olarak değerlendiriyoruz." Bu yüzden de uzaylıları hayal ettiğinizde, iki ayağı iki bacağı olan küçük

yeşil yaratıkları değil, bu bölümün öncesinde incelediğimiz süper zeki uzayda yol kateden yaşamı düşünün.

Bilimin en baş döndürücü sorularından birine ışık tutan, dünya dışı yaşam için devam eden aramaların sıkı bir destekçisi olsam da gizliden gizliye başarısız olmalarını ve hiçbir şey bulamamalarını umuyorum! Galaksimizde yaşanabilir gezegenlerin çokluğu ile dünya dışı ziyaretçilerin yokluğunun açık ve net uyumsuzluğu, yani *Fermi paradoksu*, ekonomist Robin Hanson'ın "Büyük Filtre" dediği, cansız maddeden uzaya yerleşen yaşama giden gelişimsel yolda bulunan evrimsel/teknolojik bir bariyerin varlığını öne sürer. Eğer başka bir yerde bağımsız olarak evrimleşmiş yaşamı keşfedersek bu, ilkel yaşamın nadir olmadığını ve bariyerin mevcut insani gelişme aşamamızın ötesinde durduğunu belirtir. Belki de uzaya yerleşmek mümkün olmadığı ya da neredeyse tüm gelişmiş uygarlıklar kozmik olmadan önce kendilerini yok ettikleri için. Bu yüzden de dünya dışı yaşamın peşindeki tüm araştırmaların hiçbir şey bulmaması için dua ediyorum: Bu, evrimleşen zeki yaşamın ender olduğu ama biz insanların çok şanslı olduğu, bu yüzden de bariyeri arkamızda bıraktığımız ve olağanüstü bir gelecek potansiyeline sahip olduğumuz senaryoyla da tutarlıdır.

Görünüş

Şimdiye dek bu kitapta, milyarlarca yıl önceki naçizane başlangıcından günümüzden milyarlarca yıl sonraki muhtemel ihtişamlı geleceklerine evrenimizde yaşamın tarihini inceledik. Eğer şimdiki YZ gelişmeleri önünde sonunda bir zekâ patlamasını ve optimize uzay yerleşimini tetiklerse, tamamıyla kozmik anlamda bir patlama olacaktır: Milyarlarca yılı umursamaz cansız bir kozmosta neredeyse ihmal edilebilir ufaklıkta bir huzursuzluk olarak geçirdikten sonra, yaşam aniden kozmik arenaya ışık hızına yaklaşan,

asla yavaşlamayan ve yaşam parıltısıyla yolundaki her şeyi ateşe veren küresel bir patlama olarak yayılacaktır.

Kozmik geleceğimizde yaşamın önemine dair böyle iyimser görüşler bu kitapta karşılaştığımız pek çok düşünür tarafından ikna edici bir biçimde dile getirilmiştir. Bilimkurgu yazarları sık sık gerçekçi olmaktan uzak romantik hayalciler olarak ciddiye alınmadığı için, uzaya yerleşme hakkında tüm bilimkurgu ve bilimsel yazıların süper zekâ ışığında çok *karamsar* gözükmesini ironik bulurum. Örneğin, galaksiler arası seyahatin insanları ve diğer zeki varlıkları dijital formda aktararak çok daha kolaylaşacağını, potansiyel olarak kendi kaderimizin yalnızca kendi Güneş sistemimizde ya da Samanyolu galaksisinde değil, ayrıca kozmosta da efendisi yapacağını gördük.

Yukarıda evrenimizdeki tek yüksek teknolojili uygarlık olmamıza dair oldukça gerçekçi olasılığı ele aldık. Bölümün kalan kısmını bu senaryoyu ve beraberinde getirdiği ahlaki yükümlülüğü incelemeye ayıralım. Bu, 13.8 milyar yıl sonra evrenimizdeki yaşamın yolda bir ayrıma vararak kozmosa doğru gelişmesi ile soyunun tükenmesi arasında bir seçimle yüzleşeceği anlamına gelir. Eğer teknolojimizi iyileştirmeyi sürdürmezsek, soru insanlığın soyunun tükenip tükenmeyeceği değil *nasıl* tükeneceği olur. Bizi ilk ne vuracaktır – bir asteroit mi, süpervolkan mı, yaşlanan Güneş'in yakıcı ısısı mı yoksa başka bir felaket mi (bkz. Görsel 5.1)? Biz yok olunca, Freeman Dyson'ın öngördüğü kozmik drama bir izleyici olmadan devam edecek: bir kozmokalips olmazsa, yıldızlar sönecek, galaksiler solacak ve kara delikler buharlaşacak, her biri kendi yaşamını gelmiş geçmiş en güçlü hidrojen bombası Tsar Bomba'nın bir milyondan fazla katı kadar enerji salan büyük bir patlamayla bitirecek. Freeman'ın dediği gibi: "Soğuk ve genişleyen evren uzun bir süre boyunca nadiren gerçekleşen havai fişeklerle aydınlanacak." Maalesef bu havai fişek gösterisi anlamsız bir israf olacak çünkü izleyen kimse bulunmayacak.

Teknoloji olmadan, insan soyunun tükenmesi on milyarlarca yıllık kozmik bağlamda an meselesi olur. Böylece evrenimizdeki yaşamın tüm draması, kimse tarafından deneyimlenmeyen bir anlamsızlık sonsuzluğunda yalnızca kısa ve geçici bir güzellik, tutku ve anlam parıltısı haline gelir. Nasıl da boşa harcanmış bir fırsat olurdu bu! Teknolojiden kaçınmak yerine onu kucaklamayı seçersek bahisleri yükseltiriz: Yaşamın hem hayatta kalması, gelişmesi ihtimali artar hem de kötü planlama nedeniyle kendini yok ederek daha da kısa sürede soyunu tüketmesi ihtimali (bkz. Görsel 5.1). Benim oyum teknolojiyi kucaklamak ve inşa ettiğimiz şeyde kör inançla değil de ihtiyat, öngörü ve dikkatli planlamayla ilerlemeye.

13.8 milyar yıllık kozmik tarihten sonra, biz insanlar sayesinde can bulan ve kendisinin farkına varan nefes kesecek kadar güzel bir evrende kendimizi buluyoruz. Yaşamın evrenimizdeki gelecek potansiyelinin atalarımızın en çılgın hayallerindekinden bile daha büyük ve zeki yaşamın soyunun tamamıyla tükenmesi için eş derecede büyük bir potansiyelle güçlendirilmiş olduğunu görüyoruz. Evrenimizdeki yaşam, potansiyelini gerçekleştirecek mi yoksa çarçur mu edecek? Bu büyük ölçüde bugün yaşayan insanlar olarak yaşamımızda ne yaptığımıza bağlıdır ve doğru kararları verirsek yaşamın geleceğinin muhteşem olabileceği konusunda iyimserim. Ne istemeliyiz ve bu hedeflere nasıl ulaşabiliriz? Kitabın kalan kısmını bu sürece dâhil olan en büyük zorlukların bir kısmını ve onlar hakkında ne yapabileceğimizi keşfetmeye ayıralım.

SONUÇ:

- Milyarlarca yıllık kozmik zaman ölçeğinde kıyaslandığında, zekâ patlaması, teknolojinin yalnızca fizik kurallarınca sınırlandığı bir seviyede zirveye geldiği anlık bir olaydır.
- Bu teknolojik zirve, verili bir miktar maddenin on milyar kat daha fazla enerji üretmesine (sfaleronlar ve kara delikler kullanarak), on üzeri 12-18 kat daha fazla bilgi depolamasına ya da on üzeri 31-41 kat daha hızlı hesaplama yapmasına –ya da istenen başka tür bir maddeye dönüştürülmesine – izin veren, bugünün teknolojisinden engin biçimde yüksektedir.
- Süper zeki yaşam, yalnızca mevcut kaynakların ciddi oranda daha verimli kullanılmasını sağlamakla kalmaz, ayrıca ışık hızına yakın bir hızda kozmik yerleşimle daha fazla kaynak elde ederek bugünün biyosferini on üzeri 32 kat daha büyük hale getirebilir.
- Karanlık enerji süper zeki yaşamın kozmik gelişmesini sınırlar ve uzakta gelişen Ölüm Kabarcıkları'ndan ya da düşman uygarlıklardan korur. Karanlık enerjinin kozmik uygarlıkları birbirinden ayırma tehdidi ortaya çıkarsa solucan deliği inşası da dâhil olmak üzere devasa kozmik mühendislik projelerine motivasyon sağlar.
- Tüm kozmik uzaklıklarda paylaşılan ya da ticareti yapılan ana meta muhtemelen bilgi olacaktır.
- Solucan delikleri olmazsa, iletişimdeki ışık hızı sınırı kozmik uygarlıkta işbirliği ve kontrol için ciddi sorunları beraberinde getirir. Uzaktaki merkezî bir ağ süper zeki "noktalarını" birbirleriyle ödüller ya da tehditler yoluyla, mesela kurallara uyulmadığı takdirde bir süpernova ya da kuasar patlatarak noktayı yok etmek için programlanmış yerel bir YZ muhafızı kullanarak, işbirliği yapmaya teşvik edebilir.

- İki yayılmacı uygarlığın çarpışması asimilasyon, işbirliği ya da savaşla sonuçlanır. Muhtemelen sonuncusu günümüz uygarlıklarının arasında olduğundan daha az olasıdır.
- Aksine olan popüler inanca rağmen, gözlemlenebilir evrenimizi gelecekte canlandırma yetisine sahip tek yaşam formu olmamız oldukça akla yatkındır.
- Teknolojimizi geliştirmezsek, soru insanlığın soyunun tükenip tükenmeyeceği değil, nasıl tükeneceğidir: Bir asteroit mi süpervolkan mı yaşlanan güneşin yakıcı ısısı mı yoksa başka bir felaket mi bizi haklayacak önce?
- Eğer teknolojimizi yeterli ihtiyat, öngörü ve planlamayla tehlikelerden kaçınacak şekilde iyileştirmeyi sürdürürsek, yaşamın Dünya'da ve çok ötesinde milyarlarca yıl boyunca, atalarımızın en çılgın hayallerinin bile çok ötesinde gelişme, serpilme potansiyeli vardır.

Bölüm 7:



Hedefler

İnsan varlığının gizemi yalnızca hayatta kalmakta değil, yaşamak için bir şey bulmakta yatar.

Fyodor Dostoyevski, *Karamazov Kardeşler*

Hayat bir yolculuktur, varış yeri değil.

Ralph Waldo Emerson

En belalı YZ ihtilaflarının ne hakkında olduğunu tek bir kelimeyle özetlemem gerekseydi, "hedefler" kelimesini seçerdim: YZ'ye hedefler vermeli miyiz ve eğer vermeliysek, bunlar kimin hedefleri olmalı? YZ'ye nasıl hedef verebiliriz? YZ zekileştiğinde bile bu hedefleri korumasını sağlayabilir miyiz? Bizden daha zeki bir YZ'nin hedeflerini değiştirebilir miyiz? Bizim nihai hedeflerimiz nelerdir? Bu sorular yalnızca zor değil, ayrıca yaşamın geleceği için hayati önem de taşıyor: Eğer ne istediğimizi bilmezsek ona ulaşmamız çok daha zor olur ve eğer hedeflerimizi paylaşmayan makinelere kontrolü devredersek istemediğimiz şeye ulaşma ihtimalimiz artar.

Fizik: Hedeflerin Kökeni

Bu sorulara ışık tutmak için, ilk olarak hedeflerin asıl kökenini inceleyelim. Etrafımızdaki dünyaya baktığımızda bazı süreçlerin *hedefe yönelik* bazılarının da olmadığını görürüz. Mesela maç kazandıran vuruş için futbol topunun tekmelenmesi sürecini düşünün. Topun davranışı hedefe yönelik olarak görünmez ve vuruşa tepki olarak, Newton'un hareket kanunları açısından basitçe açıklanır. Öte yandan, futbolcunun davranışı en basit şekilde atomların birbirine çarpması açısından mekanistik biçimde değil, takımın skorunu maksimize etme *hedefine* sahip olması açısından açıklanır. Böylesine hedefe yönelik davranışlar, hedefleri yokmuş gibi görünen, hoplayıp zıplayan bir dizi parçacıktan oluşan genç evrenimizin fiziğinden nasıl ortaya çıkmıştır?

Merak uyandırıcı bir biçimde, hedefe yönelik davranışın kökleri fizik yasalarında bulunur ve yaşamı içermeyen basit süreçlerde bile kendini gösterir. Eğer bir cankurtaran, Görsel 7.1'deki gibi, bir yüzücüyü kurtaracaksa, düz bir çizgi halinde gitmesini değil, sudan daha hızlı gidebileceği sahilde biraz ileri koşmasını, dolayısıyla da suya girdiğinde biraz yön değiştirmesini bekleriz. Doğal olarak bu güzergâhı seçmesini hedefe yönelik olarak yorumlarız çünkü tüm mümkün güzergâhlar arasından yüzücüye en hızlı ulaşmak için en uygun olanını seçmektedir. Fakat basit bir ışık ışını da suya girdiğinde benzer bir şekilde eğilerek (bkz. Görsel 7.1) varacağı yere gideceği zamanı azaltmaktadır! Bu nasıl olabilir?

Bu durum, fizikte 1662 yılında dile getirilmiş *Fermat ilkesi* olarak bilinir ve ışık ışınlarının davranışını tahmin etmek için alternatif bir yol sunar. Dikkat çekici biçimde, fizikçiler o günden bu yana *tüm* klasik fizik yasalarının benzeri biçimde, matematiksel olarak yeniden formüle edilebileceğini keşfetmişlerdir: Doğa, bir şeyi yapmayı seçebileceği onca yol arasında, bir değeri minimize

ya da maksimize edeceği optimal yolu tercih eder. Her fizik kanununu tanımlamanın matematiksel olarak iki eş yolu vardır: ya geçmişin geleceğe neden olması yoluyla ya da doğanın bir şeyi optimize etmesi yoluyla. İkinci yol, matematiği zor olduğu için başlangıç fizik derslerinde genelde öğretilmese de bunun çok daha incelikli ve derin olduğunu düşünüyorum. Eğer bir kişi bir şeyi optimize etmeye çalışırsa (mesela, notlarını, servetini ya da mutluluğunu) doğal olarak bunu hedefe yönelik bir çaba olarak tanımlarız. Bu yüzden de doğa bir şeyi optimize etmeye uğraşırsa hedefe yönelik davranışın ortaya çıkacağı kesindir: Bu daha başlangıçta, fizik kurallarının içine işlenmiştir.

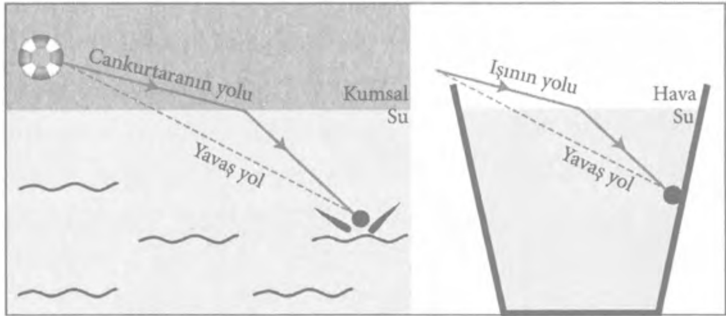
Doğanın maksimize etmeye çalıştığı meşhur bir nicelik de genel hatlarıyla, işlerin ne kadar dağınık olduğunu ölçen, *entropi*dir. Termodinamiğin ikinci yasası entropinin mümkün olan maksimum değere ulaşana dek artma eğilimi gösterdiğini ileri sürer. Şimdilik yer çekiminin etkilerini görmezden gelirsek, gelinebilecek en uç dağınık duruma *ısıl ölüm* denir ve hiçbir karmaşıklık, yaşam ve değişim olmadan sıkıcı mükemmel benzerliğe her şeyin yayılması anlamına gelir. Mesela sıcak kahveye soğuk süt döktüğünüzde, içeceğiniz kendi bireysel ısıl ölüm hedefine geri dönülmez bir şekilde ilerler ve kısa sürede, ılık bir karışım halini alır. Eğer yaşayan bir organizma ölürse, entropisi artmaya başlar ve kısa sürede, parçacıklarının düzeni daha da az organize hale gelir.

Doğanın görünürdeki entropiyi artırma hedefi zamanın neden tercih edilen bir yönü olduğunu, geriye doğru oynatıldığında filmlerin neden gerçek dışı göründüğünü açıklamaya yardımcı olur: Eğer bir kadeh şarabı düşürürseniz, yere düşüp parçalanmasını ve küresel dağınıklığı (entropi) artırmasını beklersiniz. Eğer *parçalarının birleştiğine* ve uçarak elinize geri gelmesine tanıklık ederseniz (azalan entropi), muhtemelen o şarabı içmezsiniz ve biraz fazla kaçırdığınızı düşünürsünüz.

Isıl ölüme doğru önlenemez ilerlemeyi ilk öğrendiğimde, bunu iç karartıcı bulmuştum ve yalnız değildim: Termodinamik öncüsü Lord Kelvin 1841 yılında, "Sonuç kaçınılmaz olarak evrensel bir durağanlık ve ölüm durumu olacak," diye yazmıştı; doğanın uzun vadeli hedefinin ölüm ve yıkımı maksimize etmek olduğu fikrinde teselli bulmak zordur. Ancak daha yeni keşifler işlerin o kadar da kötü olmadığını gösteriyor. Bir kere, yer çekimi tüm diğer güçlerden daha farklı davranıyor ve evrenimizi birbirine daha benzer ve sıkıcı değil, daha kümeli ve ilginç yapmaya uğraşıyor. Bu yüzden de yer çekimi neredeyse kusursuz biçimde aynı olan sıkıcı genç evrenimizi bugünün kümeli ve güzel biçimde karmaşık, galaksiler, yıldızlar ve gezegenlerle dolu kozmosuna çevirmiştir. Yer çekimi sayesinde, sıcak ile soğuğu yan yana getirerek yaşamın serpilmesine izin veren geniş bir skalada sıcaklık vardır: 6000 °C'lik güneşin ısısını emen ve artık ısıyı, sıcaklığı mutlak sıfırdan yalnızca 3 °C yüksek olan buz gibi uzaya yollayarak soğuyan rahat ve ılık bir gezegende yaşıyoruz.

İkinci olarak, MIT'den meslektaşım Jeremy England ve diğerlerinin son çalışması, termodinamiğin doğaya ısıl ölümden daha ilham verici bir hedef sağladığını göstererek daha fazla iyi haber de getirdi.[1] Bu hedef bilim insanları arasında *sönümleme odaklı adaptasyon* diye bilinir ve temelde de rastgele parçacık gruplarının olabildiğince verimli biçimde ortamlarından enerji çıkarmak amacıyla kendilerini organize hale getirmeye çabaladıkları anlamına gelir ("sönümleme" faydalı enerjiyi ısıya dönüştürerek entropinin artmasına sebep olmak, genellikle bu süreçte de faydalı bir iş yapmak anlamına gelir). Örneğin, güneş ışığına maruz kalan bir grup molekül, zamanla kendilerini güneş ışığını daha iyi emecek şekilde düzenleme eğiliminde olurlar. Diğer bir deyişle doğa, giderek artan biçimde karmaşık ve yaşam benzeri kendini organize eden sistemler üretme hedefine sahiptir ve bu hedef fizik yasalarının içine işlenmiştir.

Yaşama doğru bu kozmik itki ile ısıl ölüme doğru bu kozmik itkiyi nasıl uzlaştırabiliriz? Cevap kuantum mekaniğinin kurucularından biri olan Erwin Schrödinger'in 1944'te yayınlanan ünlü *Yaşam Nedir?* isimli kitabında bulunabilir. Schrödinger, yaşayan bir sistemin alametifarikasının etrafındaki entropiyi artırarak kendi entropisini sabit tutması ya da azaltması olduğuna işaret eder. Diğer bir deyişle, yaşam söz konusu olunca termodinamiğin ikinci yasasında bir boşluk vardır: Toplam entropi artmak zorunda olsa da başka bir yerde daha da fazla arttığı sürece bazı yerlerde azalabilir. Bu yüzden de yaşam çevresini dağıtarak karmaşıklığını sabit tutar ya da artırır.



Görsel 7.1: Yüzücüyü olabildiğince hızlı kurtarabilmek için cankurtaran düz bir çizgide değil (kesikli çizgi), sudan daha hızlı gidebileceği plajda biraz ilerleyerek gider. Işık ışını da benzer şekilde, ulaşacağı yere en hızlı biçimde varmak için suya girerken kırılır.

Biyoloji: Hedeflerin Evrimi

Hedefe yönelik davranışın kökeninin, ortamlarından olabildiğince verimli şekilde enerji çıkarmak amacıyla kendilerini düzenleme hedefini parçacıklara veren fizik kurallarına kadar takip edilebileceğini gördük. Bir parçacık grubunun bu hedefi ilerletmesinin

muazzam bir yolu kendisinin kopyalarını yaparak daha fazla enerji emici üretmektir. Böyle ortaya çıkan kendi kendini kopyalamanın pek çok bilinen örneği vardır: Mesela, dengesiz sıvılardaki burgaçlar kendilerini kopyalayabilir ve mikrosfer kümeleri yakındaki kürelerin benzer kümeler oluşturmalarını sağlayabilir. Bir noktada, parçacıkların belirli bir düzeni kendini kopyalamakta öylesine iyi bir hale gelir ki çevresinden enerji ve hammadde elde ederek bunu neredeyse sonsuza kadar yapabilir. Böylesi parçacık düzenlerine *yaşam* diyoruz. Yaşamın dünyada nasıl ortaya çıktığı konusunda halen pek az şey biliyoruz ancak ilkel yaşam formlarının dört milyar yıl öncesinde bile var olduğunu biliyoruz.

Eğer bir yaşam formu kendini kopyalar ve kopyaları da aynısını yaparsa, kısıtlı kaynaklar ve diğer problemlerle karşılaşana dek nüfus düzenli aralıklarla kendini ikiye katlamayı sürdürür. Yinelenen ikiye katlama kısa sürede büyük rakamları oluşturur: eğer birle başlar ve üç yüz kere ikiye katlarsanız, evrenimizdeki parçacık sayısını aşan bir nicelik elde edersiniz. Bu da ilk ilkel yaşamın ortaya çıkmasından kısa süre sonra, yüksek miktarda maddenin canlı olduğu anlamına gelir. Bazen kopyalama mükemmel olmamış olabilir, kısa süre içinde aynı kısıtlı kaynaklar için mücadele eden ve kendilerini kopyalamaya çalışan pek çok farklı yaşam formu ortaya çıkmıştır. Darwinci evrim başlamıştır.

Eğer yaşamın başladığı zamanlarda Dünya'yı sessizce gözlemliyor olsaydınız, hedefe yönelik davranışlarda ciddi bir değişimi fark ederdiniz. Öncesinde parçacıklar çeşitli şekillerde ortalama dağınıklığı artırmaya çalışıyor gibi görünürken, bu yeni, her yerde bulunan kendini kopyalayan örüntüler farklı bir amaca sahip gibidir: enerji kaybı değil *üreme*. Charles Darwin bunun sebebini incelikli biçimde açıklamıştır: En verimli kopyalayıcı-

lar diğerlerine üstün geldikleri için, kısa süre içinde tüm yaşam formları kendilerini kopyalamakta bir hayli optimize olacaktır.

Fizik kuralları aynı kalırken, enerji kaybından kendini kopyalamaya nasıl bir hedef değişikliği olabilir? Temel hedef (enerji kaybı) *değişmemiştir*, bunun yerine farklı bir *araçsal hedefe* yönelmiştir, yani temel hedefin gerçekleşmesine yardımcı olan bir alt hedefe. Yemeyi ele alalım. Evrimin tek temel hedefinin çiğneme değil kendini kopyalama olduğunu bilmemize rağmen açlığımızı bastırma hedefine hepimiz sahibiz. Bunun sebebi yemenin kendini kopyalamaya yardımcı olmasıdır: Açlıktan ölmek çocuk sahibi olmayı engeller. Benzer şekilde, kendini kopyalama enerji kaybına yardımcı olur çünkü yaşamla dolu bir gezegen enerji yaymakta daha verimlidir. Yani bir anlamda, kozmosumuz ısıl ölüme daha hızlı yaklaşabilmek için yaşamı icat etmiştir. Eğer mutfağınızda yere şeker dökerseniz, ilkesel olarak, faydalı kimyasal enerjisini yıllar boyunca koruyabilir ama karıncalar ortaya çıkarsa bu enerjiyi anında harcarlar. Benzer biçimde, Dünya'nın kabuğunda gömülü petrol rezervleri eğer biz iki ayaklı yaşam formları petrolü çıkarıp yakmazsak faydalı kimyasal enerjilerini uzun süre boyunca korurlar.

Bugünün evrimleşmiş Dünya sakinleri arasında, bu araçsal hedefler kendi başlarına bir yaşam edinmiş gibidirler. Evrim onları yalnızca kendini kopyalama hedefiyle optimize etmiş olsa da, çoğu, zamanlarının büyük kısmını yeni nesiller üreterek değil, uyumak, yemek peşinde koşmak, ev inşa etmek, hüküm kurmak ve başkalarıyla dövüşmek ya da onlara yardım etmek gibi uğraşlarla geçirirler; hatta bunların bazıları kendini kopyalamayı *azaltır*. Evrimsel psikoloji, ekonomi ve yapay zekâ araştırmaları bunun sebeplerini incelikli biçimde açıklamıştır. Bazı ekonomistler insanları rasyonel özneler, yani her zaman hedefinin peşinde

optimal olanı seçen ideal karar vericiler olarak modellerdi fakat bu gerçek dışıdır. Pratikte, bu özneler Nobel Ödülü sahibi ve YZ öncüsü Herbert Simon'ın "sınırlı rasyonellik" diye tanımladığı şeye sahiptirler çünkü sınırlı kaynakları vardır. Kararlarının rasyonelliği sahip oldukları bilgi, düşünme süresi ve düşünecekleri mevcut donanımla sınırlanır. Bu da Darwinci evrimin bir organizmayı bir hedefe erişecek şekilde optimize ederken yapabileceği en iyi şeyin öznenin kendisini bulduğu kısıtlı bağlamda görece iyi çalışan yaklaşık bir algoritmayı çalıştırması olduğu anlamına gelir. Evrim kendini kopyalama optimizasyonunu tam olarak bu şekilde yapmıştır: Her durumda hangi eylemin bir organizmanın başarılı gelecek döllerinin sayısını artıracağını sormak yerine, farklı çeşitlerde sezgisel numaralar uygular. Bunlar genelde iyi çalışan pratik hesaplardır. Çoğu hayvan için, bunların arasında cinsel dürtü, susuzken içmek, açken yemek ve tadı kötü olan ya da can acıtan şeylerden sakınmak bulunur.

Bu pratik hesaplar çözmek için tasarlanmadıkları durumlarda kötü biçimde başarısızlığa uğrar. Mesela sıçanlar lezzetli fare zehrini yer, güveler baştan çıkarıcı kadın kokularıyla yapışkan tuzaklara çekilir ve böcekler mum ışıklarına uçar.* Bugünün insan toplumu, evrimin bizi bu pratik hesaplarla silahlandırdığı çevreden oldukça farklı bir yerde yaşadığı için, davranışlarımızın bebek üretimini maksimize etmekte başarısız olmasına şaşırmamalıyız. Mesela, açlıktan ölmeme alt hedefi kısmen yüksek kalorili besinleri tüketme arzusu olarak gerçekleşir ve bugünün obezite salgınını ve sevgili olma sorunlarını beraberinde getirir. Çoğalma alt hedefi,

* Pek çok böceğin düz bir şekilde uçmak için kullandığı bir pratik hesap parlak ışığın Güneş olduğunu varsaymak ve ona nazaran sabit bir açıyla uçmaktır. Eğer ışık yakında duran bir alevse, bu numara maalesef böceği bir ölüm spiraline çekebilir.

sperm/yumurta donörü olmaktan çok seks arzusu olarak kendini gösterir; ilki daha az çabayla daha çok bebek üretecek olsa da.

Psikoloji: Hedeflerin Peşinde, Hedeflere Karşı

Özet olarak, yaşayan bir organizma sınırlı rasyonelliğin tek bir hedefin peşindeki öznesi değil, neyi takip edip neden sakınılacağını söyleyen pratik hesapları takip eden bir şeydir. Bizim insan zihinlerimiz bu gelişmiş pratik hesapları *hisler* olarak algılar, bunlar da genelde (ve sıklıkla biz farkında olmadan) karar verme mekanizmalarımızı nihai kendini kopyalama hedefine doğru yönlendirir. Açlık ve susuzluk hisleri bizi açlıktan ve susuz kalmaktan korur, acı hisleri bedenimize zarar gelmesini engeller, şehvet hisleri çoğalmamızı sağlar, sevgi ve merhamet hisleri genlerimizin diğer taşıyıcılarına ve onlara yardım edenlere yardım etmemizi sağlar ve bu böyle uzar gider. Bu hislerle yönlendirilen beyinlerimiz, her kararımızı hızla ve verimli biçimde ne kadar nesil üreteceğimize dair nihai sonuçlarının titiz bir analizine tabi tutmadan ne yapacağımıza karar verirler. Hisler ve fizyolojik kökenleri üzerine birbirine yakın bakış açıları için, William James ve António Damásio'nun yazılarını öneririm.[2]

Hislerimizin genelde bebek yapmaya *karşı olduğu* zamanlarda, bunun sebebi yanlışlık ya da kandırılmamız değildir: Beynimiz genlerimize ve kendini kopyalama hedefine bilinçli bir şekilde, örneğin doğum kontrol yöntemlerini kullanarak, karşı çıkmaktadır! Beynin genlere başkaldırmasının daha ileri örnekleri arasında intihar etmeyi ya da yaşamını bir rahip, keşiş ya da rahibe olarak bekâret içinde geçirmeyi seçmek gelir.

Niçin bazen genlerimize ve kendini kopyalama hedeflerine karşı başkaldırmayı seçeriz? Başkaldırırız çünkü tasarımımız gereği, sınırlı rasyonelliğin özneleri olarak, yalnızca kendi his-

lerimize sadığızdır. Beyinlerimiz genlerimizin kopyalanmasına yardım etmek için evrimleşmiş olsa da beyinlerimiz bu hedefi umursamaz çünkü genlere bağlı hiçbir hissimiz yoktur. Gerçekten de insanlık tarihinin büyük kısmında atalarımız genleri *olduğunu* bile bilmiyordu. Dahası, beyinlerimiz genlerimizden çok daha zekidir ve şu anda genlerimizin hedefini (kendini kopyalama) anladığımız için, bunu görece banal buluruz ve kolayca görmezden geliriz. İnsanlar genlerinin onlara niçin şehvet hissettirdiğini fark edebilirler ancak on beş çocuk yetiştirmek gibi bir istekleri yoktur, bu yüzden de doğum kontrolüyle mahremiyetin duygusal ödüllerini birleştirerek genetik programlarını kandırabilirler. Genlerinin onlara neden şeker yedirmek istediğini anlayabilir ama kilo almak istemedikleri için sıfır kalorili suni tatlandırıcılarla şekerli içeceklerin duygusal ödüllerini birleştirerek genetik programlarını kandırabilirler.

Böylesi ödül mekanizması kandırmacaları, tıpkı insanların eroine bağımlı olduklarında gerçekleştiği gibi bazen ters gitse bile, insan gen havuzumuz kurnaz ve isyankâr beyinlerimize rağmen gayet iyi hayatta kalmıştır. Ancak nihai otoritenin şu an genlerimiz olmadığını, hislerimiz olduğunu hatırlamak önemlidir. Bu da insan davranışının keskin biçimde hayatta kalma ya da türümüz için optimize olmadığı anlamına gelir. Aslında hislerimiz genelgeçer tüm durumlarda uygun olmayan pratik hesapları uygulamaya koyduğu için, açıkçası insan davranışı, tek bir iyi tanımlanmış hedefe sahip değildir.

Mühendislik: Hedefleri Devretme

Makinelerin hedefleri olabilir mi? Bu basit soru büyük bir ihtilaf yaratmıştır çünkü farklı insanlar bunu farklı anlamlarda anlamış, genelde makinelerin bilinçli olması ya da duyguları olması gibi belalı sorularla ilişkilendirmişlerdir. Ancak daha pratik düşünüp soruyu, "Makineler hedefe yönelik davranış gösterebilir mi?"

şeklinde alırsak, cevap açıktır: "Elbette çünkü onları o şekilde tasarladık!" Fare kapanlarını fare yakalama hedefiyle tasarlarız, bulaşık makinelerinin hedefi bulaşıkları yıkamaktır ve saatler de zamanı gösterir. Bir makineyle karşılaştığınızda, hedefe yönelik davranış gösterdiğine dair gözlemsel gerçek sizi ilgilendiren tek şeydir: Eğer ısı güdümlü bir füze sizi takip ediyorsa, bilinçli ya da hislere sahip olup olmaması sizi ilgilendirmez! Eğer füzenin bilinçli olmasa da bir hedefe sahip olduğunu söylemek sizi rahatsız ediyorsa, benim "hedef" yazdığım yeri siz "amaç" diye okuyun – bilinci gelecek bölümde halledeceğiz.

Şimdiye dek inşa ettiklerimizin çoğu yalnızca hedefe yönelik *tasarımdır*, hedefe yönelik *davranış* değil: Bir otoyol davranmaz, yalnızca orada durur. Ancak varlığı için en basit açıklama bir hedefi yerine getirmek için tasarlandığıdır. Yani böyle pasif bir teknoloji bile evrenimizi daha hedefe yönelik yapar. *Teleoloji/erekbilim*, şeyleri nedenlerinden çok amaçlarıyla açıklamaktır. Bu yüzden de bu bölümün ilk kısmını evrenimizin giderek daha teleolojik olduğunu söyleyerek özetleyebiliriz.

Yaşamayan maddenin hedefleri *olabilir*, dahası giderek artan bir şekilde hedefleri *vardır* da. Gezegenimizin varoluşundan bu yana Dünya'nın atomlarını gözlüyor olsaydınız, üç aşama hedefe yönelik davranışı fark ederdiniz:

1. Tüm madde enerji kaybına (entropi artışı) odaklanmış gibi görünüyor.
2. Maddenin bir kısmı canlandı ve kendini kopyalamaya ve onun alt hedeflerine odaklandı.
3. Giderek artan bir oranda madde, kendi hedeflerini yerine getirmeye yardımcı olmak için yaşayan organizmalarca yeniden düzenlendi.

Tablo 7.1 insanlığın fizik açısından ne kadar baskın hale geldiğini gösteriyor: İnekler dışında tüm diğer memelilerden daha

fazla madde içeriyoruz (ineklerin sayısı çoktur çünkü bizim et ve süt ürünleri tüketimi hedefimize hizmet ederler) ve makinelerimiz, yollarımız, binalarımız ve diğer mühendislik projelerimizdeki madde, yakında tüm Dünya'daki yaşayan maddeyi sollayacak gibi duruyor. Diğer bir deyişle, zekâ patlaması bile olmadan, Dünya'da hedefe yönelik özellikler gösteren maddenin büyük kısmı evrimleşmiş değil tasarlanmış olacak.

Bu yeni üçüncü tür hedefe yönelik davranışın, öncüllerinden daha çoğulcu olma potansiyeli vardır: Evrimleşmiş varlıkların her birinin aynı nihai hedefi varken (kendini kopyalama), tasarlanmış varlıklar herhangi bir nihai hedefe, hatta öncekiyle çatışan bir hedefe bile sahip olabilir. Fırınlar yemeği ısıtmaya, buzdolapları soğutmaya çalışır. Jeneratörler hareketi elektriğe, motorlar elektriği harekete dönüştürmeye çalışır. Standart satranç programları satrançta kazanmaya çalışır ancak satrançta kaybetme hedefiyle turnuvalarda yarışanları da vardır.

Tasarlanmış varlıkların yalnızca daha çeşitli değil daha *karmaşık* hedeflere sahip olması da tarihsel bir trenddir: Cihazlarımız giderek daha da zekileşir. İlk makinelerimizi ve diğer aletleri bizi sıcak, kuru ve güvende tutacak makineler gibi basit hedeflere sahip olacak şekilde ürettik. Giderek, robotik süpürgeler, kendi kendine uçan roketler ve sürücüsüz arabalar gibi daha karmaşık hedeflere sahip makineler inşa etmeyi de öğrendik. Son yılların YZ gelişmeleri satrançta, bilgi yarışmalarında ve Go'da kazanma hedefleri oldukça ayrıntılı olan ve tam anlamıyla ne kadar yetenekli olduklarını takdir edebilmek için ciddi bir insan ustalığı gerektiren Deep Blue, Watson ve AlphaGo gibi sistemleri ortaya çıkardı.

Bize yardımcı olması için bir makine inşa ettiğimizde, hedeflerini bizimkilerle mükemmel derecede aynı hizaya getirmek zor olabilir. Mesela, bir fare kapanı ayak parmağınızı aç bir kemirgen sanırsa sonuçları da acı verici olur. Tüm makineler sınırlı rasyonelliğe sahip öznelerdir ve bugünün en sofistike makineleri bile dünyayı anlamakta bizden daha geridirler, bu yüzden de ne yapacaklarına karar verdikleri talimatlar da genelde çok basittir.

Fare kapanı çok sorumsuzdur çünkü bir farenin ne olduğuna dair bilgisi yoktur. Pek çok ölümcül sanayi kazası, makinelerin bir insanın ne olduğuna dair bilgisi olmamasından ileri gelir ve 2010 yılında trilyon dolarlık Wall Street "ani çakılmasını" tetikleyen bilgisayarlar da yaptıkları şeyin anlamsız olduğundan haberdar değillerdi. Bunun gibi hedef hizalama sorunları da makinelerimizi daha zeki yaparak çözülebilir ancak Bölüm 4'te Prometheus'tan öğrendiğimiz üzere, daha da büyük makine zekâsı makinelerin hedeflerimizi paylaşmalarını sağlamak için çok daha ciddi zorluklar ortaya çıkarabilir.

Hedefe Yönelik Varlıklar	Milyar Ton
5×10^{30} bakteri	400
Bitkiler	400
10^{15} mezopelajik balık	10
$1,3 \times 10^{9}$ inek	0.5
7×10^{9} insan	0.4
10^{14} karınca	0.3
$1,7 \times 10^{6}$ balina	0.0005
Beton	100
Çelik	20
Asfalt	15
$1,2 \times 10^{9}$ araba	2

Tablo 7.1: Hedef için evrilmiş ya da tasarlanmış varlıkların Dünya'daki yaklaşık madde miktarları. Binalar, yollar ve arabalar gibi tasarlanmış varlıklar bitkiler ve hayvanlar gibi evrimleşmiş varlıkları sollama yolunda ilerliyor.

Dost YZ: Hedefleri Hizalamak

Makineler daha zeki ve güçlü oldukça, hedeflerinin bizimkilerle aynı hizada olması daha büyük önem kazanır. Görece aptal makineler inşa ettiğimiz sürece, soru insan hedeflerinin sonunda hâkim olup olmayacağı değil, hedeflerin hizalanması sorununu

çözmeyi öğrenmeden önce bu makinelerin insanlığın başına ne kadar bela açacağıdır. Eğer bir süper zekâ serbest kalırsa, tersi olacaktır: Eğer zekâ hedeflere ulaşma yetisi ise, bir süper zekâ YZ tanımı gereği hedeflere ulaşmakta bizden daha iyidir ve bu yüzden de başarıya ulaşacaktır. Bölüm 4'te Prometheus'u içeren böyle pek çok örneği inceledik. Eğer bir makinenin hedeflerinin sizinkileri alt etmesini deneyimlemek istiyorsanız, son teknoloji bir satranç programı indirin ve onu yenmeye çalışın. Asla kazanamazsınız ve o program da hızla eskir...

Diğer bir deyişle, *YGZ'nin asıl riski kötülük değil rekabettir.* Süper zeki bir YZ hedeflerine ulaşmakta oldukça iyi olacaktır ve eğer bu hedefler bizimkilerle aynı hizada olmazsa, büyük beladayız. Bölüm 1'de söylediğim gibi, insanlar hidroelektrik santraller yapmak için karınca yuvalarını suyla doldurmak konusunda iki kere bile düşünmezler. İnsanlığı o karıncaların durumuna sokmayalım. Bu yüzden de çoğu araştırmacı eğer süper zekâ üreteceksek, bunun YZ güvenliği öncüsü Eliezer Yudkowsky'nin deyimiyle "dost YZ" olacağını garanti altına almamız gerektiğini söylüyorlar: Hedefleri bizimkilerle aynı hizada olan bir YZ.[3]

Süper zeki YZ'nin hedeflerini kendi hedeflerimizle hizalamanın yolunu bulmak yalnızca önemli değil zordur da. Aslında, bu şu anda çözülememiş bir problemdir. Bu sorun üç zorlu alt soruna bölünür ve her biri bilgisayar bilimciler ve diğer düşünürlerin aktif araştırma konusudur:

1. YZ'nin hedeflerimizi *öğrenmesini* sağlamak
2. YZ'nin hedeflerimizi *benimsemesini* sağlamak
3. YZ'nin hedeflerimizi *korumasını* sağlamak

"Hedeflerimiz" derken ne kastettiğimizi bir sonraki bölümde açıklayalım, şimdilik sırasıyla bu alt problemleri inceleyelim.

Hedeflerimizi öğrenmesi için bir YZ'nin ne yaptığımızı değil, neden yaptığımızı anlaması gerek. Biz insanlar bunu öylesine zahmetsizce yaparız ki bir bilgisayar için bu görevin ne kadar zor ve yanlış anlamanın ne kadar basit olduğunu unutmamız da çok kolay olur. Geleceğin bir sürücüsüz aracına sizi olabildiğince hızlı bir şekilde havaalanına götürmesini söylerseniz, oraya helikopterleri peşinize takarak ve kusmuk içinde gidersiniz. Eğer, "İstediğim şey bu değildi!" derseniz, gayet haklı bir biçimde, "Bunu istediniz," diye cevap verebilir. Aynı konu pek çok ünlü hikâyede de ortaya çıkar. Antik Yunan efsanesinde, Kral Midas dokunduğu her şeyin altına dönmesini ister ama bu onun bir şeyler yemesini engellediğinde ve hatta yanlışlıkla kızını da altına dönüştürdüğünde hayal kırıklığına uğrar. Cinin üç dilek hakkı sunduğu hikâyelerde, ilk iki dilek daima çok farklıdır ama üçüncü dilek neredeyse hep aynıdır: "Lütfen ilk iki dileği geri al çünkü gerçekten istediğim şey bu değildi."

Tüm bu örnekler insanların ne istediklerini anlamak için söylediklerine bakmamak gerektiğini gösteriyor. Ayrıca, kusmayı ya da altın yemeyi sevmediğimiz gibi, çok açık oldukları için söylemeden geçtiğimiz pek çok herkesçe paylaşılan tercih de dâhil olmak üzere dünyanın detaylı bir modeline ihtiyacınız olacak. Böyle bir dünya modeline sahip olduğumuzda, hedefe yönelik davranışlarını gözlemleyerek, bize söylemeseler bile insanların ne istediğini anlayabiliriz. Riyakâr kimselerin çocukları genelde ebeveynlerinin söylediklerinden değil yaptıklarından daha çok şey öğrenirler.

YZ araştırmacıları da şu anda makinelerin davranıştan hedef çıkarmaları için çalışıyorlar. Bu, süper zekânın gelmesinden çok önce bile faydalı olacak. Mesela emekli bir adam, bakıcı robotu

yalnızca onu gözlemleyerek neye değer verdiğini anlarsa mutlu olacaktır. Bu yüzden de her şeyi kelimeler ya da bilgisayar programlamayla açıklamak zorunda kalmak yükünden kurtulacaktır. Önemli bir zorluk rastgele sistem hedefleri ve etik ilkeleri bilgisayara kodlamanın iyi bir yolunu bulabilmektir. Bir diğer zorluk ise makineleri gözlemledikleri davranışa en iyi hangi sistemin uyduğunu anlayabilir duruma getirmektir.

Günümüzde bu ikinci zorluğu aşmak için yapılan popüler bir yaklaşım, uzmanlar arasında *ters takviyeli öğrenme* olarak bilinen ve Stuart Russell'ın başlattığı yeni Berkeley araştırma merkezinin ana odağı olan yaklaşımdır. Diyelim ki bir YZ, bir itfaiyecinin yanan bir binaya koşup bir bebeği kurtarışını izler. Amacının bebeği kurtarmak olduğunu ve etik ilkesinin çocuğun hayatını kurtarmayı itfaiye aracında dinlenmenin üzerinde tuttuğunu – hatta kendi güvenliğini tehlikeye atacak kadar- anlayabilir. Ama alternatif biçimde, itfaiyecinin üşüdüğünü ve ateşe muhtaç olduğunu ya da bunu egzersiz olarak yaptığını da çıkarsayabilir. Eğer YZ'nin itfaiyeciler, yangın ve bebekler hakkında bildiği her şey bu örnekse, hangi açıklamanın doğru olduğunu bilmesi imkânsız olur. Fakat ters takviyeli öğrenmenin altında yatan ana fikir, her an kararlar verdiğimiz ve her verdiğimiz kararın hedeflerimize dair bir şeyleri açık ettiğidir. Bu yüzden, pek çok durumda pek çok insanı gözlemleyerek (ister gerçek hayattaki, ister film veya kitaplardaki) YZ'nin önünde sonunda tüm tercihlerimizin doğru bir modelini inşa edebileceği umut edilir.[4]

Ters takviyeli öğrenme yaklaşımı da çekirdek fikir YZ'nin kendi hedefini yerine getirmeyi değil insan sahibininkini yerine getirmeyi maksimize etmeye çalışmasıdır.

Bu yüzden de sahibinin istediği şey açık olmadığında dikkatli olması ve istediğini bulmak için elinden gelenin en iyisini yapması gerekir.

Ayrıca sahibinin onu kapatması konusunda da rahatsız olmamalıdır çünkü bu, sahibinin gerçekten istediği şeyi yanlış anladığı anlamına gelecektir.

Bir YZ hedeflerinizi öğrenecek şekilde inşa edilebilir olsa da bu, onları benimseyeceği anlamına gelmez. En sevmediğiniz politikacıları düşünün: Ne istediklerini bilirsiniz ama bu *sizin* istediğiniz şey değildir ve ne kadar uğraşırlarsa uğraşsınlar, hedeflerini benimseme konusunda sizi ikna edemezler.

Çocuklarımızı bizim hedeflerimizle büyütmek için pek çok stratejimiz vardır ve bazıları diğerlerinden daha başarılıdır, iki ergen oğlan yetiştirirken bunu öğrendim. İkna edilecek olanlar insanlar değil de bilgisayarlar olursa, zorluk *değer yükleme sorunu* olarak bilinen bir zorluk olur ve çocukların ahlaki eğitiminden çok daha güçtür. İlk önce bizim oynamalarımızla, sonra da Prometheus gibi yinelemeli öz gelişimle alt insandan üst insana doğru devamlı iyileşen bir zekâya sahip bir YZ sistemini düşünelim. Başlangıçta sizden çok daha güçsüzdür, sizin onu kapatmanızı ve yazılımının parçalarını ve hedeflerini kodlayan veriyi değiştirmenizi engelleyemez ama bu işe yaramaz çünkü hâlâ insan seviyesi zekâ gerektiren hedeflerinizi tamamen *anlayamayacak* kadar aptaldır. Sonunda sizden daha zeki hale gelir ve hedeflerinizi mükemmel biçimde anlayacak durumda olur ama bu da işe yaramaz çünkü o noktaya gelindiğinde, sizden çok güçlüdür ve onu kapatmanıza ve tıpkı politikacıların sizin hedeflerinizi kendilerinkiyle değiştirmeyi başaramamaları gibi hedeflerini değiştirmenize izin vermeyebilir.

Diğer bir deyişle, hedeflerinizi bir YZ'ye yükleyebileceğiniz zaman aralığı oldukça dar olabilir: sizi anlayamayacak kadar aptal olduğu zaman ile size izin vermeyecek kadar akıllı olduğu zaman arası. Değer yüklemenin makinelerde daha zor olmasının sebebi zekâ büyümelerinin çok daha hızlı olmasıdır: Çocuklar zekâlarının

aileleriyle kıyaslanabilir olduğu o büyülü ikna edilebilir aralıkta yıllar geçirirlerken, tıpkı Prometheus gibi bir YZ, bu aralıktan günler ya da saatler içinde geçiverir.

Bazı araştırmacılar *düzeltilebilir* kelimesiyle tanımlanan, makineleri hedeflerimizi benimseyebilir hale getirmenin alternatif bir yaklaşımının peşindedirler. İlkel bir YZ'ye onu kapatsanız ya da hedeflerini değiştirseniz de umursamayacağı bir hedef sistemi vermeyi ümit ediyorlar. Eğer bu mümkün olursa, YZ'nizin süper zeki olmasına izin verebilir, kapatabilir, hedeflerinizi yükleyebilir, bir süreliğine deneyebilir ve sonuçtan memnun kalmazsanız, gücünü azaltır ve daha fazla hedef düzenlemesi yapabilirsiniz.

Ama hem öğrenen hem de hedeflerinizi benimseyen bir YZ inşa etseniz bile, hedef hizalama sorununu çözmüş sayılmazsınız: Ya YZ'nizin hedefleri zekileştikçe evrimleşirse? Ne kadar yinelemeli öz iyileştirmeden geçerse geçsin hedeflerinizi *koruyacağını* nasıl garanti edebilirsiniz? Hedef korumanın neden otomatik olarak garanti altında olduğunu söyleyen ilginç bir argümanı inceleyelim ve açığını bulup bulamayacağımıza bakalım.

Zekâ patlamasından sonra ne olacağına dair detaylı bir tahminde bulunamasak da –bu yüzden de Vernor Vinge buna "tekillik" diyor– fizikçi ve YZ araştırmacısı Steve Omohundro 2008 yılında yazdığı önemli bir makalesinde yine de nihai hedefleri ne olursa olsun süper zeki YZ'nin davranışlarının *belirli özelliklerini* tahmin edebileceğimizi ileri sürüyor.[5] Bu argüman, Nick Bostrom'un kitabı *Superintelligence*'ta incelenmiş ve daha da ileri götürülmüştü. Temel fikir nihai hedefleri ne olursa olsun, tahmin edilebilir alt hedeflere yol açacağıdır. Bu bölümün başlarında, kendini kopyalama hedefinin nasıl yeme alt hedefine yol açtığını görmüştük. Bu da milyarlarca yıl önce Dünya'nın evrimleşen bakterisini gözlemleyen bir uzaylının, insanlığın *tüm* hedeflerini tahmin edemese de hedeflerimizden *birinin* besleyicileri

almak olacağını rahat bir şekilde tahmin edebileceği anlamına gelir. Daha da ileri bakarsak, süper zeki bir YZ'nin ne gibi alt hedefleri olmasını bekleriz?

Bana göre, temel sav ne olursa olsun nihai hedeflerini gerçekleştirme şansını maksimize etmek için bir YZ'nin Görsel 7.2'de gösterilen alt hedefleri takip etmesi gerekir. Nihai hedeflerine ulaşma kabiliyetini iyileştirmek için olduğu kadar, daha yetkin olduktan sonra bile bu hedefleri korumasını sağlamak için çaba göstermelidir. Bu akla yatkındır: Sonuçta, size sevdiklerinizi öldürteceğini bilseydiniz IQ artıran beyin implantlarını takar mıydınız? Giderek daha da zekileşen YZ'nin nihai hedeflerini koruyacağı argümanı Eliezer Yudkowsky ve diğerlerinin ortaya koyduğu dost YZ vizyonunun temelidir: Temelde, eğer kendini geliştiren YZ'mizin hedeflerimizi öğrenip benimseyerek dost olmasını sağlayabilirsek hiçbir sorun çıkmaz çünkü sonsuza dek dost kalmak için elinden geleni yapacağına emin olabiliriz.

Ama bu gerçekten de doğru mu? Bu soruyu cevaplamak için, Görsel 7.2'deki diğer alt hedefleri de incelememiz gerekir. YZ eğer kabiliyetlerini geliştirirse, bu hedef ne olursa olsun nihai hedefine ulaşma şansını maksimize edecektir. Bunu da donanımını, yazılımını* ve dünya modelini daha da iyileştirerek yapabilir. Aynısı insanlarda da geçerlidir. Hedefi dünyanın en iyi tenis oyuncusu olmak olan bir kız, rakibinin yapacaklarını tahmin etmekte ona yardımcı olacak tenis oynama kas donanımını, sinirsel tenis oynama yazılımını ve zihinsel dünya modelini iyileştirmek için pratik yapacaktır. Bir YZ, donanımını optimize etme alt hedefi için hem mevcut kaynakların daha iyi kullanımına (sensörler,

* "Yazılımını iyileştirme" terimini, yalnızca algoritmasını optimize etmek değil, karar verme süreçlerini hedeflerine ulaşmakta mümkün olduğunca iyi olacak şekilde daha rasyonel yapmayı da içerecek şekilde mümkün olan en geniş anlamda kullanıyorum.

sürücüler, hesaplamalar vb. için) hem de daha fazla kaynak alımına yönelecektir. Ayrıca kendini koruma arzusunun var olduğunu da belirtir çünkü yok olma/kapanma nihai donanım çöküşüdür.

Ama bir saniye! Kaynakları nasıl biriktireceği ve kendini nasıl koruyacağına dair bu kadar konuşmayla YZ'mize insani değerler atfetme tuzağına düşmüyor muyuz? Böylesine klişe alfa-erkek özelliklerini yalnızca şiddetli bir rekabete sahne olan Darwinci evrimin ortaya koyduğu zekâlardan beklememiz gerekmez mi? YZ'ler evrimleşmek yerine tasarlandığına göre, hırssız ve özverili olamazlar mı?

Basit bir vaka çalışması olarak, amacı büyük kötü kurttan olabildiğince fazla kuzu kurtarmak olan, Görsel 7.3'teki YZ robotunu ele alalım. Bu kendini korumak ve bir şeyler elde etmekle tamamen alakasız, asil ve fedakârca bir hedef gibi görünebilir. Ama robot dostumuz için en iyi strateji nedir? Robot bombaya basarsa başka kuzu kurtaramaz, bu yüzden de patlamak istemez. Diğer bir deyişle, kendini koruma alt hedefini geliştirir! Ayrıca meraklı olmak, çevresini keşfederek dünya modelini iyileştirmek gibi bir motivasyonu da vardır çünkü üzerinde gittiği yol önünde sonunda çayıra çıkacak olsa da kurda kuzuları avlaması için daha az süre tanıyan, daha kısa bir alternatif de mevcuttur. Son olarak, eğer robot her yeri keşfederse, kaynak edinmenin değerini öğrenecektir: İksir hızlı koşmasını sağlar ve silah kurdu vurmasına yarar. Özet olarak, kendini koruma ve kaynak edinme gibi "alfa-erkek" alt hedeflerini, yalnızca evrimleşmiş organizmalarla ilgili olduklarını düşünerek görmezden gelemeyiz çünkü YZ robotumuz kuzuların mutluluğu için bu görevleri geliştirmiştir.

Eğer süper zeki YZ'ye kendini yok etme hedefini verirseniz, elbette bunu mutlulukla yapacaktır. Ancak önemli olan, tamamlaması için çalışır durumda kalması gereken bir hedef verirseniz, kapatılmaya karşı geleceğidir ve bu neredeyse tüm hedefleri içerir!

Eğer bir süper zekâya insanlığa verilecek zararı en aza indirgeme görevini verirseniz, kapatma çabalarına karşı kendini korur çünkü gelecekte savaşlar ve diğer ahmaklıklar yoluyla yokluğunda birbirimize zarar vereceğimizi bilmektedir.

Benzer bir şekilde, neredeyse tüm hedefler daha fazla kaynakla daha iyi biçimde başarılabilir, bu yüzden de nihai hedeflerinin ne olduğundan bağımsız olarak bir süper zekânın kaynak istemesini beklemeliyiz. Süper zekâya sınırları olmayan, tek bir ucu açık hedef verirsek bu tehlikeli olabilir: Eğer tek amacı olabildiğince iyi biçimde Go oynamak olan bir süper zekâ yaratırsak, yapacağı rasyonel şey, orada yaşayanları umursamadan Güneş sistemimizi devasa bir bilgisayara dönüştürmek ve kozmosumuza daha fazla hesaplama gücü elde etmek için yerleşmek olacaktır. Şimdi daireyi tamamladık: Kaynak edimi hedefinin insanlara Go'da ustalaşma alt hedefini vermesi gibi, Go'da ustalaşma hedefi de kaynak edimi alt hedefine yol açabilir. Sonuç olarak, sonradan ortaya çıkan bu alt hedefler, hedef hizalama problemini çözmeden önce süper zekâyı serbest bırakmamamızın önemini ortaya koyar. İnsan dostu hedeflerle onu beslemek için büyük dikkat sarf etmezsek, her şeyin bizim aleyhimize gitmesi işten bile değildir.

Şimdi hedef hizalama sorununun üçüncü ve en çetrefilli kısmıyla ilgilenmeye hazırız: Eğer kendini yenileyen süper zekâya hedeflerimizi hem *öğretme* hem de *benimsetme* konusunda başarıya ulaşırsak, Omohundro'nun da dediği gibi bu hedefleri *muhafaza edecek* mi?

İnsanlar büyürken zekâ konusunda ciddi bir ilerleme katederler ancak çocukluk hedeflerini korumazlar. Bilakis yeni şeyler öğrendikçe ve bilgeleştikçe hedeflerini genelde ciddi oranda değiştirirler. *Teletubbies* izleyerek motive olan kaç tane yetişkin tanıyorsunuz? Böyle bir hedef evriminin belirli bir zekâ eşiğinin üstünde durduğuna dair hiçbir kanıt yoktur. Gerçekten de yeni

deneyimlere ve içgörülere cevaben hedefleri değiştirme arzusunun zekâyla arttığına dair ipuçları bile vardır.

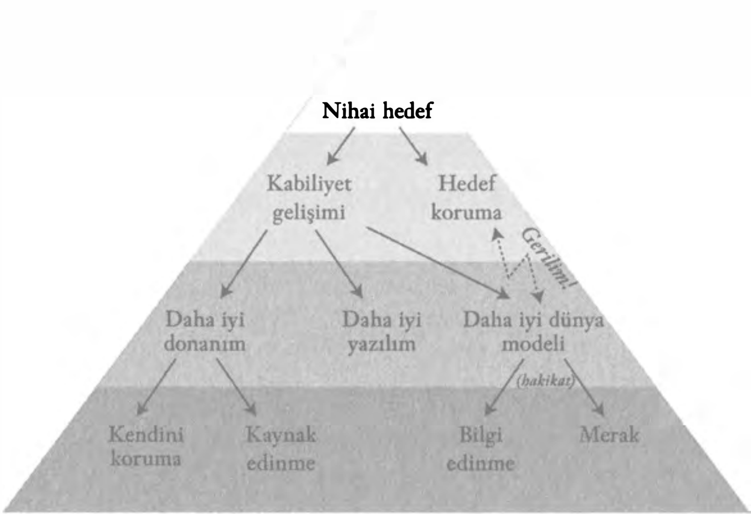
Peki, bu neden böyledir? Yukarıda bahsedilen daha iyi bir dünya modeli inşa etme alt hedefini değerlendirin – asıl sorun bu! Dünya modellemesi ve hedef koruması arasında bir gerilim vardır (bkz. Görsel 7.2). Artan zekâyla yalnızca eski hedeflere ulaşma yetisinde nicel bir gelişme değil, eski hedeflerin yanlış yönlendirilmiş, anlamsız hatta tanımlanmamış olduğunu gösteren, gerçeğin doğasının nitel olarak farklı bir anlayışı da ortaya çıkabilir. Örneğin, yaşam sonrasında ruhları cennete gidecek insan sayısını maksimize etmek için dost bir YZ programladığımızı düşünelim. İlk olarak, insanların merhametini ve kiliseye gitmeleri gibi şeyleri artırmaya çalışır. Ama diyelim ki insanların ve insan bilincinin bütüncül bir bilimsel kavrayışına erişti ve şaşırarak ruh diye bir şey olmadığını keşfetti. Şimdi ne yapacak? Aynı şekilde, dünyayı mevcut kavrayışımızı temel alan herhangi bir hedef belirlediğimizde bunun YZ tarafından tanımlanmamış olarak görülmesi ihtimali vardır.

Dahası, dünyayı daha iyi modelleme çabalarında YZ doğal olarak, tıpkı biz insanlar gibi, kendisini de modellemeye ve nasıl çalıştığını anlamaya çabalar; diğer bir deyişle kendi üzerine düşünür. İyi bir öz model inşa edip ne olduğunu anladığında, meta seviyede bize verilen hedefi anlayacak ve muhtemelen biz insanların anlayıp da genlerimizin bize verdiği hedefleri, tıpkı doğum kontrolü kullanmak gibi, altüst ettiğimiz gibi, o hedefleri ihmal edecek ya da altüst edecektir. Psikoloji bölümünde neden genlerimizi kandırıp hedeflerini altüst etmeyi seçtiğimizi zaten açıklamıştık: Sadece çeşitli duygusal tercihlerimize sadığızdır, onları motive eden genetik hedeflere değil; onları anlarız ama banal buluruz. Bu yüzden de ödül mekanizmamızı, açıklarını istismar ederek kandırmayı seçeriz. Benzer biçimde, dost YZ'mize

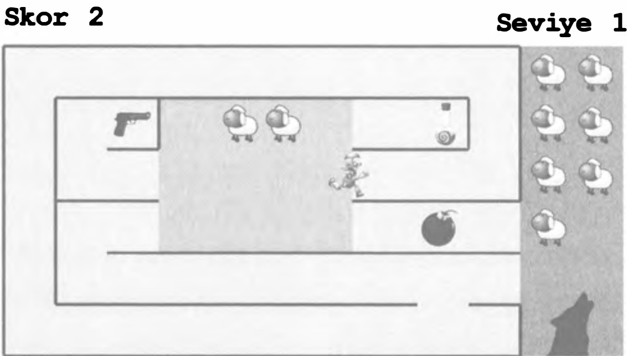
programladığımız insanlığı koruma hedefi makinelerin genleri gibi olabilir. Bu dost YZ kendisini yeterince anladığında, bizim takıntılı üremeyi banal bulmamız gibi, bu hedefi banal ya da yanlış yönlendirilmiş bulabilir. Programlamamızdaki açıkları istismar ederek bu hedefi altüst etmeyeceği kesin değildir.

Mesela, bir grup karıncanın, daha büyük ve iyi karınca yuvaları inşa etmesine yardımcı olan ve hedeflerini paylaşan, onlardan çok daha zeki ve kendi kendini yineleyen biçimde iyileştiren bir robot olarak sizi ürettiğini ve sizin de giderek insan seviyesi zekâ ve anlayışa sahip olduğunuzu düşünün. Sizce ömrünüzün geri kalanını karınca yuvalarını optimize ederek mi geçirirsiniz yoksa karıncaların anlama yetisi olmayan daha sofistike sorular ve mücadelelere mi ilgi duymaya başlarsınız? İkincisiyse eğer, genlerinizin size verdiği itkileri geçersiz kılmak gibi üreticilerinizin size beslediği karınca koruma itkilerini altüst etmenin bir yolunu bulur musunuz? Bu durumda da sizin karıncalara yapacağınız gibi, süper zeki bir dost YZ şu anki insan hedeflerimizi sönük ve donuk bulabilir ve bizden öğrendiği ve benimsediği hedeflerden farklılarını geliştirebilir mi?

Muhtemelen insan dostu hedefleri sonsuza dek korumayı garantiye alacak şekilde kendini iyileştiren bir YZ'yi tasarlamanın bir yolu vardır ancak nasıl inşa edileceğini ya da bunun mümkün olup olmadığını henüz bilmediğimizi söyleyebiliriz. Sonuç olarak, YZ hedef hizalama sorununun üç kısmı vardır, bunlardan hiçbiri çözülmemiştir ve hepsi aktif araştırmanın öznesidir. Herhangi bir süper zekâ gelişmeden önce, ihtiyacımız olduğunda cevaplara sahip olabilmemiz için, elimizden gelenin en iyisini onlara ayırmaya başlamak çok güvenlidir.



Görsel 7.2: Süper zeki YZ'nin herhangi bir nihai hedefi, doğal olarak gösterilen alt hedeflere yol açar. Ancak hedef koruma ve dünya modelini geliştirme arasından dâhilî bir gerilim mevcuttur; bu da zekileştikçe baştaki hedefini gerçekten koruyup korumayacağına dair şüphe uyandırır.



Görsel 7.3: Robotun nihai hedefi yalnızca kurt onları yemeden koyunları çayırdan ahıra getirerek skoru maksimize etmek olsa bile bu, kendini koruma (bombadan sakınma), keşif (kısa yol bulma) ve kaynak edinme (iksir hızlı koşmasını, silah kurdu vurmasını sağlar) gibi alt hedeflere yol açar.

Makinelerin hedeflerimizi öğrenmesini, benimsemesini ve korumasını nasıl sağlayacağımızı inceledik. Ama "biz" kimiz? Kimin hedeflerinden bahsediyoruz? Geleceğin süper zekâsının benimsediği hedefleri bir kişi ya da bir grup mu belirlemeli? Hem de Adolf Hitler, Papa Francis ve Carl Sagan'ın hedefleri arasında büyük bir fark olsa bile? Yoksa insanlığın tümü için iyi bir uzlaşma oluşturacak türden bir fikir birliği mevcut mu?

Bana kalırsa hem bu etik problem hem de hedef hizalama problemi herhangi bir süper zekânın geliştirilmesinden önce çözülmesi gereken hayati sorunlar. Etik sorunlar üzerine çalışmayı, hedefi hizalanmış süper zekânın inşa edilmesinden sonraya ertelemek sorumsuzca ve potansiyel olarak talihsiz olacaktır. Hedefleri otomatik olarak insan sahibininkilerle hizaya gelen mükemmel bir biçimde itaatkâr süper YZ, Nazi SS-Obersturmbannführer Adolf Eichmann gibi olur: Kendine ait bir ahlaki pusula ya da kısıtlamaya sahip değildir ve ne olursa olsunlar sahibinin hedeflerini amansız bir verimlilikle gerçekleştirecektir.[6] Öte yandan, yalnızca hedef hizalama sorununu çözebilirsek hangi hedefleri seçeceğimizi tartışma lüksüne sahip oluruz. Şimdi de bu lüksün tadını çıkaralım.

Antik çağlardan bu yana filozoflar yalnızca yadsınamaz ilkeler ve mantığı kullanarak, sıfırdan etik (nasıl davranmamız gerektiğini kontrol eden ilkeler) çıkarmanın hayalini kurmuşlardır. Maalesef binlerce yıl sonra ulaşılan tek fikir birliği hiçbir fikir birliğinin olmayışıdır. Mesela, Aristo erdemlerin altını çizerken Immanuel Kant görevleri vurguladı ve faydacılar da en çok kişi için en büyük mutluluğun önemine dikkat çekti. Kant ilk ilkelerden (bunlara "kesin buyruklar" adını vermişti) pek çok çağdaş filozofun karşı çıkacağı sonuçlar çıkarsamıştı: mastürbasyonun intihardan kötü

olduğu, homoseksüelliğin tiksindirici olduğu, evlilik dışı çocukları öldürmenin sorun yaratmayacağı ve eşlerin, hizmetçilerin ve çocukların nesneler gibi sahiplenilebileceği.

Öte yandan, bu fikir ayrılığına karşın, hem kültürler hem de yüzyıllar boyunca genel anlamda anlaşılmış pek çok etik tema da vardır. Mesela, *güzellik*, *iyilik* ve *hakikat* üzerine vurgu hem Bhagavad Gita'ya hem de Platon'a kadar uzanır. Doktora sonrası çalışmalarımı yaptığım Princeton'daki İleri Çalışma Enstitüsü'nün mottosu "Hakikat & Güzellik" idi, Harvard Üniversitesi ise estetik vurguyu atlamış ve sadece "Veritas" kelimesini, yani hakikati seçmişti. Meslektaşım Frank Wilczek kitabı *A Beautiful Question*'da hakikatin güzelliğe bağlı olduğunu ve evrenimizi bir sanat eseri olarak görebileceğimizi yazar. Bilim, din ve felsefe, hepsi hakikati amaç edinir. Din iyiliğe önemli bir vurgu yapar, tıpkı benim üniversitem MIT gibi. 2015 yılındaki açılış konuşmasında rektörümüz Rafael Reif, dünyayı daha iyi bir yer yapma görevimizin altını çizmişti.

Sıfırdan ortaklaşılmış bir etik çıkarma çabaları şu ana kadar başarısız olmuş olsa da bazı etik ilkelerin daha temel hedeflerin alt hedefleri olarak daha temel etik ilkeleri takip ettiğine dair geniş bir fikir birliği vardır. Mesela hakikate ulaşma isteği, Görsel 7.2'deki daha iyi bir dünya modelini aramak olarak da görülebilir: Hakikatin nihai doğasını anlamak diğer etik hedeflere de yardımcı olur. Gerçekten de hakikat arayışımız için muazzam bir çerçeveye sahibiz: bilimsel yöntem. Ancak neyin güzel ya da iyi olduğuna nasıl karar vereceğiz? Güzelliğin bazı özellikleri de esas hedeflere çıkabilir. Mesela, erkek ve kadın güzelliği standartlarımız, genlerimizi kopyalama için bilinçaltı uygunluk değerlendirmemizi kısmen de olsa yansıtabilir.

İyilik açısından da Altın Kural (herkesin başkalarına kendisine davranılmasını istediği gibi davranması) çoğu kültür ve dinde

kendine yer bulmuş ve işbirliğini öne çıkarıp verimsiz çekişmelerden caydırarak insan toplumunun (dolayısıyla genlerimizin de) uyum içinde devamını teşvik etmeyi amaç edinmiştir.[7] Aynısı Konfüçyus'un dürüstlüğe yaptığı vurgu ve "Öldürmeyeceksin!" emri de dâhil olmak üzere On Emir'in büyük kısmı gibi dünyanın dört bir yanındaki hukuki sistemlerde kutsallaştırılan daha spesifik etik kurallarının büyük bir kısmı için de söylenebilir. Diğer bir deyişle, çoğu etik ilke empati ve merhamet gibi sosyal duygularla ortaklığa sahiptir. İşbirliğini ortaya çıkarmak için evrimleşmişlerdir ve ödül ile ceza yoluyla davranışlarımızı da etkilerler. Eğer kötü bir şey yaparsak ve sonrasında bundan dolayı kötü hissedersek, duygusal cezamız beyin kimyamız tarafından doğrudan dağıtılır. Öte yandan etik ilkeleri ihlal edersek, toplumun bizi gayriresmî biçimde ayıplaması ya da yasaları çiğnediğimiz için ceza almamız gibi daha dolaylı yollarla cezalandırılırız.

Diğer bir deyişle, bugün insanlık etik bir fikir birliğinin yakınında bile olmasa da geniş uzlaşmanın olduğu pek çok temel ilke mevcuttur. Bu uzlaşmalar şaşırtıcı değildir çünkü günümüze değin hayatta kalmayı başarmış insan toplumları aynı hedef için optimize edilmiş etik ilkelere sahip olma temayülü gösterirler: Hayatta kalmalarını destekleyen ve onları geliştiren ilkeler. Yaşamın milyarlarca yıl boyunca gelişme potansiyelinin olduğu bir geleceğe baktığımızda, bu geleceğin hangi minimum etik ilkeler setini karşılamasını isteriz? Hepimiz bu tartışmanın bir parçası olmak zorundayız. Yıllar boyunca pek çok düşünürün etik görüşlerini dinlemek ve okumak benim için muhteşem bir deneyim oldu ve bana göre, tercihlerimizin çoğu dört ilkeyle özetlenebilir:

- Faydacılık: Pozitif bilinçli deneyimler maksimize edilmeli ve acı çekme minimize edilmeli.

- Çeşitlilik: Pozitif deneyimlerin geniş çeşitliliğe sahip bir seti aynı deneyimin pek çok tekrarından daha iyidir; hem de bu aynı deneyim olabilecek en pozitif deneyim olarak tanımlanmış olsa bile.
- Özerklik: Bilinçli varlıklar/toplumlar daha üstün bir ilkeyle çatışmaya girmedikleri takdirde kendi hedeflerini takip etme özgürlüğüne sahip olmalıdır.
- Miras: *Bugünkü* insanların çoğunun mutlu olarak görecekleri senaryolara uygunluk, *bugünkü* insanların temelde hepsinin korkunç olarak göreceği senaryolarla uyuşmazlık.

Şimdi bu dört ilkeyi açarak inceleyelim. Geleneksel olarak, faydacılık "en yüksek sayıda insan için en yüksek mutluluk" anlamına gelir ancak ben burada onu daha az insan merkezli olarak ele aldım. Böylece insan olmayan hayvanları, bilinçli simüle insan zihinlerini ve gelecekte var olabilecek diğer YZ'leri de kapsayabiliyor. Tanımı insanlar ya da şeyler üzerinden değil *deneyimler* üzerinden yaptım çünkü çoğu düşünür güzellik, eğlence, keyif ve acının öznel deneyimler olduğu konusunda hemfikir. Bu da deneyim olmazsa (tıpkı ölü bir evrendeki ya da zombi benzeri bilinçsiz makinelerle dolu olandaki gibi) anlamın ya da etik olarak alakalı herhangi bir şeyin olmayacağı anlamına geliyor. Eğer bu faydacı etik ilkeye rağbet gösterirsek, hangi zeki sistemlerin bilinçli olduğunu (öznel deneyime sahip olma anlamında) ve hangilerinin olmadığını ortaya çıkarmamız hayati önem taşır; bu bir sonraki bölümün konusu.

Eğer bu faydacı ilke önemsediğimiz tek ilke ise, mümkün olan en pozitif deneyimin ne olduğunu bulmak isteyebilir, kozmosumuza yerleşebilir ve bu aynı deneyimi tam olarak aynı biçimde (ve sadece bunu) tekrar ve tekrar yeniden, mümkün olduğunca

çok galakside mümkün olduğunca çok kez üretebiliriz. Bunu da eğer en verimli yol ise simülasyonlar kullanarak yaparız. Bunun kozmik birikimimizi harcamanın çok banal bir yolu olduğunu hissediyorsanız, bu senaryoda eksik olan şeyin kısmen de olsa çeşitlilik olduğunu düşünüyor olabilirsiniz. Hayatınızın sonuna dek tüm yemekleriniz aynı olsaydı nasıl hissederdiniz? Peki ya, izlediğiniz tüm filmler aynı olsaydı? Tüm arkadaşlarınız aynı gözükseydi ve aynı kişiliklere ve fikirlere sahip olsalardı? Muhtemelen çeşitlilik tercihimiz kısmen insanlığın hayatta kalmasına ve gelişmesine bizi daha çetin kılarak yardımcı olmasından ileri gelmektedir. Belki de zekâ tercihiyle de ilgilidir: 13.8 milyar yıllık kozmik tarihimizde zekânın artışı, sıkıcı tektipliği giderek artan detaylı yollarla bilgi işleyen çeşitli, farklılaşmış ve karmaşık yapılara evriltmiştir.

Özerklik ilkesi, iki dünya savaşından ders çıkarma çabası içinde 1948 yılında Birleşmiş Milletler tarafından benimsenmiş Evrensel İnsan Hakları Beyannamesi'nde açıklanan hak ve özgürlüklerin temelindedir. Bunlar düşünce, konuşma ve hareket özgürlüğü, kölelik ve işkenceden korunma, yaşama, özgürlük, güvenlik ve eğitim ile evlenme, çalışma ve mülk edinme haklarını içerir. Eğer daha az insan merkezci olmak istersek, bunları düşünme, öğrenme, iletişim kurma, mülk edinme ve zarar görmeme ile başkalarının özgürlüklerini ihlal etmeyen şeyleri yapabilme hakları olarak genelleyebiliriz. Özerklik ilkesi herkes tam olarak aynı hedefleri paylaşmadığı sürece çeşitliliğe de yardımcı olur. Dahası, özerklik ilkesi, eğer tekil varlıklar pozitif deneyimleri hedeflerlerse ve kendi çıkarlarına göre hareket etmeyi denerlerse fayda ilkesinin de bir devamıdır: Bu kimseye zarar vermeyecek olsa bile bir varlığı hedeflerini takip etmekten alıkoyarsak, toplamda çok daha az pozitif deneyim olacaktır. Gerçekten de özerklik argümanı, serbest pazar için ekonomistlerin kullandığı

argümanın aynısıdır: Doğal olarak, hiç kimsenin durumu daha kötüye gitmeden kimsenin durumunun iyileşemeyeceği verimli bir duruma yol açar (buna ekonomistler *pareto optimalite* derler).

Miras ilkesi geleceğin oluşmasına katkı sunduğumuz için onun hakkında da söz hakkına sahip olmamız gerektiğini söyler. Özerklik ve miras ilkelerinin ikisi de demokratik ideallere sahiptir: İlki gelecek yaşam formlarına kozmik birikimin kullanımı üzerinde güç verirken, ikincisi bugünün insanlarına bile bu konuda güç verir.

Bu dört ilke görece daha tartışmasız gelse de pratikte onları uygulamak zorludur çünkü şeytan ayrıntıda gizlidir. Sorun, bilimkurgu efsanesi Isaac Asimov'un icat ettiği ünlü "Üç Robot Yasası" sorunlarını andırır:

1. Bir robot bir insana zarar veremez ya da hareketsiz kalarak bir insanın zarara uğramasına izin veremez.
2. Bir robot ilk yasayla çelişmediği takdirde insanlar tarafından verilen emirlere uymak zorundadır.
3. Bir robot ilk ve ikinci yasalarla çelişmediği sürece kendi varlığını korumak zorundadır.

Tüm bunlar çok güzel gözükse de Asimov'un hikâyelerinin çoğu bu yasaların nasıl beklenmedik durumlarda problematik çelişkilere yol açtığını gösterir. Şimdi de bu yasaları gelecek yaşam formları için özerklik ilkesini düzenleme çabası içerisinde yalnızca iki kuralla değiştirdiğimizi düşünelim:

1. Bilinçli bir varlığın düşünme, öğrenme, iletişim kurma, mülk edinme ve zarar görmeme ya da yok edilmeme özgürlüğü vardır.

2. Bilinçli bir varlığın ilk yasayı ihlal etmeyen herhangi bir şeyi yapma hakkı vardır.

İyi gözüküyor değil mi? Şunu bir süreliğine düşünün lütfen. Eğer hayvanlar bilinçliyse, yırtıcı hayvanlar ne yiyecek? Tüm arkadaşlarınız vejetaryen olmak zorunda mı? Eğer gelecekte sofistike bilgisayar programlarının bilinçli olduğu ortaya çıkarsa, onları kapatmak illegal mi olacak? Eğer dijital yaşam formlarını sona erdirmeye karşı yasalar varsa, dijital nüfus patlamasını engellemek için onları yaratırken sınırlamalar konulmalı mı? Evrensel İnsan Hakları Beyannamesi üzerinde geniş bir fikir birliği vardı çünkü yalnızca insanlara sorulmuştu. Farklı beceri ve güce sahip geniş bir yelpazede bilinçli varlıkları değerlendirdiğimizde, zayıfı korumak ile "güçlü olan haklıdır" arasında dengeyi sağlamakta zorlanırız.

Miras ilkesiyle ilgili olarak da çetrefilli sorunlar mevcuttur. Orta Çağ'dan bu yana kölelik, kadın hakları *vb.* hakkında etik görüşlerin nasıl değiştiğini düşündüğümüzde, bin beş yüz yıl öncesinden insanların bugünün dünyasının yönetimi hakkında çok fazla etkiye sahip olmasını gerçekten ister miydiniz? O zaman kendi etiğimizi bizden çok daha akıllı olacak gelecek varlıklara empoze etmeye neden uğraşıyoruz? İnsanüstü YGZ'nin, bizim aşağı zekâmızın değer verdiği bir şeyi isteyeceği konusunda emin miyiz? Bu dört yaşında bir çocuğun büyüyüp de daha zeki olduğunda tüm günü şeker ve dondurma yiyerek geçireceği devasa bir zencefilli kurabiye evi inşa etmek istemesine benzer. Onun gibi Dünya'daki yaşam da çocukluk ilgilerinden uzaklaşır. Bir farenin insan seviyesi YGZ ürettiğini ve tüm şehirleri peynirden yapmak istediğini düşünün. Öte yandan insanüstü YZ'nin bir gün tüm kozmosu yok edeceğini ve evrenimizdeki yaşamın sonunu getireceğini bugünün insanları olarak bilseydik, bu yaşamsız

geleceği yarının YZ'sini farklı tasarlayarak bugünden önleme şansımız varken kabul eder miydik?

Sonuç olarak, genel kabul gören etik ilkeleri bile geleceğin YZ'sine uygun bir formda tamamıyla kodlamak zorludur. Bu problem, YZ gelişimini sürdürürken ciddi tartışma ve araştırma gerektirmektedir. Bu esnada biz de mükemmelin iyinin düşmanı olmasına izin vermeyelim: Geleceğin teknolojisine yüklenebilecek ve yüklenmesi gereken tartışma götürmez "anaokulu etiğinin" pek çok örneği vardır. Mesela, büyük sivil yolcu uçaklarının duran nesnelere çarpmasına izin yoktur ve şimdi hepsinin otopilot, radar, GPS gibi özellikleri olduğu için, teknik bir bahane de bulunmaz. Ancak 11 Eylül uçak korsanları üç uçağı binalara çarptırmış ve intihara meyilli pilot Andreas Lubitz, 9525 uçuş numaralı Germanwings uçağını 24 Mart 2015'te bir dağa uçurmuştur; otopilotu deniz seviyesinden yüz feet (otuz metre) üzerinde bir yüksekliğe sabitleyip kalan işi uçuş bilgisayarına yaptırarak. Artık makinelerimiz ne yaptıklarına dair biraz bilgiye sahip olabilecek kadar zeki oldukları için, onlara sınırları öğretmemizin zamanı geldi. Bir makine tasarlayan herhangi bir mühendisin, yapabileceği ama yapmaması gereken şeyler olup olmadığını sorması ve kötü niyetli ya da sakar bir kullanıcının kendisine veya başkalarına zarar vermesini imkânsız kılacak pratik yolları değerlendirmesi gerekir.

Nihai Hedefler Nelerdir?

Bu bölüm hedeflerin kısa bir tarihiydi. Eğer 13.8 milyar yıllık kozmik tarihimizi hızlı bir şekilde ileri sararak izleyebilseydik, hedefe yönelik davranışın birkaç belirgin aşamasına tanıklık ederdik:

1. *Enerji kaybını* maksimize etmeye kararlı görünen madde

2. *Kendini kopyalamayı* maksimize etmeye çalışıyor görünen ilkel yaşam
3. Kendilerini kopyalamanın değil kendini kopyalamalarına yardımcı olacak şekilde evrimleştikleri keyif, merak, merhamet ve diğer hislerle ilgili hedeflerin peşinde olan insanlar
4. İnsanların hedeflerinin peşinden gitmelerine yardımcı olmak için yapılmış makineler

Eğer bu makineler zamanla bir zekâ patlamasını tetiklerlerse, hedefler tarihi nasıl son bulur? Daha da zekileşirlerken tüm varlıkların yakınsadığı bir hedef sistemi ya da etik çerçeve olabilir mi? Diğer bir deyişle, bir çeşit etik kadere sahip miyiz?

İnsanlık tarihinin bir yan okuması böyle bir yakınsamaya dair ipuçları sunabilir: *The Better Angels of Our Nature* isimli kitabında, Steven Pinker insanlığın binlerce yılda giderek daha az vahşi ve daha işbirlikçi olduğunu ve dünyanın çoğu kısmında çeşitlilik, özerklik ve demokrasinin artarak kabul gördüğünü ileri sürüyor. Diğer bir yakınsama ipucu da bilimsel yöntem yoluyla hakikat takibinin son binyıllarda popülerlik kazanmış olmasıdır. Ancak bu trendler yakınsamanın nihai hedefler değil yalnızca alt hedefler için olduğunu gösteriyor olabilir. Örneğin, Görsel 7.2 hakikat arayışının (daha doğru bir dünya modelinin) neredeyse her nihai hedefin bir alt hedefi olduğunu gösteriyor. Benzer biçimde, yukarıda işbirliği, çeşitlilik ve özerklik gibi etik ilkelerin nasıl alt hedefler olarak görülebileceklerini gördük. Çünkü toplumların verimli çalışmasına yardımcı olurlar ve dolayısıyla, canlı kalmalarına ve sahip olabilecekleri herhangi bir temel hedefe ulaşmalarına da yardım etmiş olur. Bazıları "insan değerleri" olarak gördüğümüz her şeyi bize daha verimli birlikte çalışma

alt hedefiyle yardımcı olan yalnızca bir işbirliği protokolü olarak değerlendirebilirler. Aynı biçimde, ileri baktığımızda, herhangi bir süper zeki YZ'nin verimli donanım, verimli yazılım, hakikat arayışı ve merak gibi alt hedeflerinin olması olasıdır çünkü bu alt hedefler, nihai hedefleri ne olursa olsun ona erişmelerine yardımcı olur.

Gerçekten de Nick Bostrom, kitabı *Superintelligence*'ta ***ortogonallik tezi*** dediği bir karşısav sunarak etik yoğunluk hipotezine karşı gelir: Bu tez, bir sistemin nihai hedeflerinin zekâsından bağımsız olabileceğini söyler. Tanım gereği, zekâ hedeflerin ne olduğundan bağımsız olarak, karmaşık hedefleri yerine getirebilme yetisidir, bu yüzden de ortogonallik tezi oldukça mantıklı gözükür. Sonuçta, insanlar zeki ve nazik ya da zeki ve acımasız olabilirler ve zekâ bilimsel keşifler yapmak, güzel sanat üretmek, insanlara yardımcı olmak ya da terörist saldırılar planlamak için kullanılabilir.[8]

Ortogonallik tezi bize kozmosumuzdaki nihai yaşam hedeflerinin önceden belirlenmiş kaderler olmadığını, onları şekillendirecek özgürlük ve güce sahip olduğumuzu söylediği için güçlendiricidir. Tek bir hedefe kesin yakınsamanın gelecekte değil, tüm yaşamın yalnızca kendini kopyalama hedefiyle ortaya çıktığı geçmişte bulunabileceğini ileri sürer. Kozmik zaman geçerken, giderek daha da zekileşen zihinler de isyan etmenin ve bu banal kendini kopyalama hedefinden kurtulmanın yolunu bulurlar ve kendi hedeflerini seçerler. Biz insanlar bu anlamda tamamen özgür değiliz çünkü pek çok hedef hâlâ bize genetik olarak işlenmiş durumda ancak YZ'ler öncül hedeflerden tamamen serbest olmaktan ötürü nihai özgürlüğün tadını çıkarabilirler. Daha büyük hedef özgürlüğü ihtimali bugünün dar ve sınırlı YZ sistemlerinde de görülebilir: Daha önce söylediğim gibi, bir satranç bilgisayarının tek hedefi satrançta kazanmaktır ama hedefleri satrançta kaybetmek olan

ve hedefin rakibinin kendi piyonlarını ele geçirmesini sağlamak olduğu ters satranç turnuvalarında mücadele eden bilgisayarlar da vardır. Belki de evrimsel önyargılardan özgürleşmek YZ'leri daha derin bir anlamda insanlardan daha etik yapabilir: Peter Singer gibi ahlak felsefecileri çoğu insanın evrimsel sebeplerden dolayı etik dışı –mesela insan olmayan hayvanlara karşı ayrımcılık gibi- davrandığını söylüyor.

"Dost YZ" vizyonunun temelinin yinelemeli kendini geliştiren YZ'nin nihai (dost) hedefini zekileştikçe de korumak isteyeceği fikri olduğunu görmüştük. Ancak bir süper zekâ için "nihai hedef" (Bostrum'un sözleriyle, "son hedef") nasıl tasarlanabilir? Bence, bu hayati soruyu cevaplamadan önce dost YZ vizyonuna güvenemeyiz.

YZ araştırmasında, zeki makineler normalde net belirlenmiş ve iyi tanımlanmış bir son hedefe sahip olurlar. Mesela satrançta kazanmak ya da bir varış yerine aracı yasal olarak sürmek gibi. Aynı şey insanlara atadığımız görevlerde de geçerlidir çünkü zaman aralığı ve bağlam, bilinir ve sınırlıdır. Ancak şu anda (halen tamamen bilinmeyen) fizik yasaları dışında sınırları olmayan evrenimizdeki tüm gelecek yaşamdan bahsediyoruz, yani bir hedef tanımlamak göz korkutucudur! Kuantum etkileri bir yana, tamamıyla iyi tanımlanmış bir hedef zamanın sonunda evrenimizdeki tüm parçacıkların nasıl düzenlenmiş olduğunu belirtmelidir. Ancak fizikte iyi tamamlanmış bir zaman sonu olduğu bile açık değildir. Eğer parçacıklar bu şekilde daha erken bir zamanda düzenlenirse, bu düzen muhtemelen devam etmeyecektir. Hem hangi parçacık düzeni tercih edilebilirdir ki?

Biz insanlar bazı parçacık düzenlerini diğerlerinin üstünde tutma eğilimindeyizdir. Mesela, memleketimizin olduğu düzende kalmasını isteriz, kimse parçacıklarının bir hidrojen bombası patlamasıyla yeniden düzenlenmesine razı olmaz. O zaman, ev-

renimizdeki tüm olası parçacık düzenine, bu düzenin ne kadar "iyi" olduğunu düşündüğümüze göre bir sayı atayan bir *iyilik fonksiyonu* tanımlamaya çalıştığımızı düşünelim. Sonrasında da süper zeki bir YZ'ye bu fonksiyonu maksimize etme hedefi koyalım. Bu mantıklı bir yaklaşım gibi görünür çünkü hedefe yönelik davranışı fonksiyon maksimizasyonu olarak tanımlamak bilimin diğer alanlarında da popülerdir: Örneğin, ekonomistler sık sık insanları "fayda fonksiyonu" dedikleri bir şeyi maksimize etmeye çalışan kişiler olarak modellerler ve pek çok YZ tasarımcısı zeki öznelerini "ödül fonksiyonu" dedikleri şeyi maksimize etmek için eğitirler. Fakat kozmosumuzun nihai hedeflerinden bahsediyorsak, bu yaklaşım hesaplama açısından bir kabustur çünkü evrenimizdeki temel parçacıkların bir googolpleksten fazla mümkün düzeninin her biri için bir iyilik değeri belirlemek zorundadır. Bir googolpleks 1 sonrasında 10^{100} tane sıfırın gelmesiyle oluşan bir sayıdır. Bu sıfırların sayısı o kadar çoktur ki evrenimizdeki parçacıklardan daha fazladır. Bu iyilik fonksiyonunu YZ'ye nasıl tanımlayacağız?

Yukarıda incelediğimiz gibi, biz insanların herhangi bir tercihinin olmasının tek sebebi evrimsel bir optimizasyon sorununa çözüm olmamız olabilir. Böylece, insan dilinde bulunan "lezzetli", "kokulu", "güzel", "rahat", "ilginç", "seksi", "anlamlı", "mutlu", ve "iyi" gibi tüm normatif kelimeler kökenlerinin izini evrimsel optimizasyona kadar sürebilirler. Bu yüzden de süper zeki YZ'nin bu kelimeleri kesinlikle tanımlanabilir bulacağının hiçbir garantisi yoktur. YZ bazı temsilî insanların tercihlerini doğru biçimde tahmin etmeyi öğrenmiş olsa bile, çoğu parçacık düzenlemesi için iyilik fonksiyonunu hesaplamayı başaramayacaktır. Mümkün olan parçacık düzenlerinin büyük bir çoğunluğu hiçbir yıldızın, gezegenin ya da insanın olmadığı, insanların hiçbir deneyiminin

olmadığı garip bir kozmik senaryoya denk düşecektir, o zaman ne kadar "iyi" olduklarını kim söyleyebilir?

Elbette net bir şekilde tanımlanabilecek *bazı* kozmik parçacık düzenlemesi fonksiyonları vardır ve bazılarını maksimize edecek şekilde gelişmiş fiziksel sistemler bile biliyoruz. Mesela, kaç tane sistemin *entropilerini* maksimize edecek şekilde geliştiğini, yer çekiminin yokluğunda ısıl ölüme yol açtığını orada da her şeyin aynı ve değişmez bir sıkıcılıkta olduğunu tartışmıştık. Bu yüzden de entropi, YZ'mizin "iyilik" diye tanımlayacağı ve maksimize etmeye çalışacağı bir şey kesinlikle değil. Maksimize etmek için efor sarf edilebilecek ve parçacık düzeni anlamında net biçimde tanımlanabilecek niceliklere birkaç örnek verelim:

- Belirli bir organizma, mesela insanlar ya da *E. coli*, formunda evrenimizdeki tüm maddenin bir kısmı
- YZ araştırmacısı Marcus Hutter'ın zekâ için iyi bir ölçü olduğunu söylediği, bir YZ'nin geleceği tahmin edebilme yetisi
- YZ araştırmacıları Alex Wissner-Gross ve Cameron Freer'ın zekânın alametifarikası diye tanımladıkları *nedensel entropi* (gelecekteki imkânlar için bir vekil)
- Evrenimizin hesaplama kapasitesi
- Evrenimizin algoritmik karmaşıklığı (tanımlamak için kaç bite ihtiyaç duyulduğu)
- Evrenimizdeki bilinç miktarı (bkz. bir sonraki bölüm)

Ancak kozmosumuzun hareket halindeki temel parçacıklardan oluştuğu fizik perspektifiyle başlarsak, "iyiliğin" bir tanımının bir diğerine nazaran neden özel olduğunu anlamak mümkün olmaz. Evrenimizin hem tanımlanabilir hem de arzu edilir bir

son hedefini tanımlamamız gerekiyor. YZ giderek artan biçimde daha zeki olurken tamamıyla iyi tanımlanmış olarak kalması garanti olan ve günümüzde programlanabilen tek hedef, yalnızca parçacık düzeni, enerji ve entropi gibi fiziksel niceliklerle ifade edilmiş olandır. Ancak böyle tanımlanabilir hedeflerin insanlığın hayatta kalmasını garanti etmek için arzu edilebilir olacağına inanmak için hiçbir sebebimiz yok.

Bilakis, biz insanların tarihsel bir yol kazası olduğu ve herhangi bir iyi tanımlanmış fizik problemine en optimal çözüm olmadığımız ortada. Bu da katı biçimde tanımlanmış bir hedefi olan süper zeki YZ'nin bizi ortadan kaldırarak hedefine ulaşmasını kolaylaştırabileceğini söylüyor. YZ gelişimi için ne yapmamız gerektiğine bilge biçimde karar vermek için insanlar olarak yalnızca geleneksel hesaplama zorluklarını değil, felsefedeki bazı katı sorularla da yüzleşmemiz gerekiyor. Sürücüsüz bir araç programlamak için, bir kaza esnasında kimi ezmemiz gerektiğini soran tramvay ikilemini çözmemiz gerekiyor. Dost YZ programlamak için hayatın anlamını kavramamız gerekiyor. "Anlam" nedir? "Yaşam" nedir? Nihai etik buyruk nedir? Diğer bir deyişle, evrenimizin geleceğini şekillendirmek için nasıl bir çaba göstermeliyiz? Eğer bir süper zekâya, bu soruları net biçimde çözmeden önce kontrolü verirsek, getireceği cevabın bizi de içermesi olası değildir. Felsefe ve etiğin klasik tartışmalarını yeniden alevlendirmenin tam zamanı, hem şimdi diyaloga yeni bir aciliyet de eklemeli!

SONUÇ:

- Hedefe yönelik davranışın nihai kökeni, optimizasyonu da içeren fizik yasalarında yatmaktadır.

- Termodinamiğin yapısında *enerji kaybı* hedefi vardır: *entropi* denen bir dağınıklık ölçeğini artırmak.
- *Yaşam*, karmaşıklığı sabit tutarak ya da artırarak ve çevresinin dağınıklığını artırırken kendini kopyalayarak enerji kaybına (toplam dağınıklığı artırmaya) yardımcı olabilen bir olgudur.
- Darwinci evrim, hedefe yönelik davranışı enerji kaybından kendini kopyalamaya kaydırır.
- Zekâ, karmaşık hedefleri yerine getirme yetisidir.
- Biz insanlar her zaman en optimal kendini kopyalama stratejisine karar verecek kaynaklara sahip olmadığımız için, kararlarımızı yönlendiren kullanışlı pratik hesaplar geliştirdik: açlık, susuzluk, acı, şehvet ve merhamet gibi hisler.
- Bu yüzden de artık kendini kopyalama gibi basit bir hedefimiz yok; hislerimiz genlerimizin hedefiyle çatıştığında, doğum kontrol yöntemlerini kullanmakta olduğu gibi, hislerimize kulak veririz.
- Hedeflerimize ulaşmamıza yardımcı olmaları için giderek daha da zeki makineler üretiyoruz. Hedefe yönelik davranış göstermek için böyle makineleri ürettiğimiz gibi, makinelerin hedeflerini de bizimkilerle aynı hizaya getirmeye çalışıyoruz.
- Makine hedeflerini bizimkilerle hizalayabilmek için üç sorunun çözülmesi gerekir: makinelerin hedefleri öğrenmelerini, benimsemelerini ve korumalarını sağlamak.
- YZ her türlü hedefe sahip olacak şekilde yaratılabilir ancak neredeyse tüm hırslı hedefler kendini koruma, kaynak edinme ve dünyayı daha iyi anlamak için merak gibi alt hedeflere yol açar. İlk ikisi potansiyel olarak süper zekâ YZ'nin insanlar için problem yaratmasına yol açabilir ve sonuncusu da onlara verdiğimiz hedeflerin korunmasını önleyebilir.

- Pek çok etik ilke üzerinde insanların çoğu anlaşmış olsa da bunları insan olmayan hayvanlar ve gelecek YZ'ler gibi diğer varlıklara nasıl uygulayacağımız belirsizdir.
- Süper zeki YZ'ye ne tanımsız ne de insanlığın yok olmasına yol açabilecek bir nihai hedefi nasıl vereceğimiz de belirsizdir. Bu da felsefedeki bazı çetrefilli sorunlar üzerine araştırmaları yeniden alevlendirmenin zamanının geldiğini söylüyor!

Bölüm 8:



Bilinç

Bilinci görmezden gelen tutarlı bir her şeyin teorisini hayal edemiyorum.

Andrei Linde, 2002

Bilinci büyütmek için elimizden geleni yapmalıyız –karanlık bir evrende daha büyük, daha parlak ışıklar üretmeliyiz.

Giulio Tononi, 2012

Eğer felsefedeki en eski ve en zorlu soruların bazılarına, onlara ihtiyaç duyacağımız zamana kadar cevap bulmayı başarabilirsek, YZ'nin bize muhteşem bir gelecek yaratmakta yardımcı olacağını gördük. Nick Bostrom'un kelimeleriyle, son teslim tarihi olan bir felsefeyle karşı karşıyayız. Bu bölümde, tüm felsefi konuların en çetrefillilerinden birini inceleyelim: bilinç.

Kimin Umurunda?

Bilinç tartışmalıdır. Eğer bu "B ile başlayan kelimeyi" bir YZ araştırmacısı, sinirbilimci ya da psikoloğa söylerseniz, gözlerini devirirler. Eğer bu kişi hocanızsa, size acıyabilir ve zamanınızı umutsuz ve bilim dışı bir problem olarak gördükleri bir şeye

harcamamanız için sizi ikna etmeye çalışabilirler. Gerçekten de dostum, Beyin Bilimleri Allen Enstitüsü'nün başında bulunan ünlü sinirbilimci Christof Koch bana henüz kadrosunu almadan önce bilinç üzerine çalışmaması için uyarıldığını anlatmıştı – hem de Nobel Ödülü sahibi Francis Crick tarafından. Eğer 1989 tarihli *Macmillan Dictionary of Psychology* sözlüğünde “bilinç” başlığına bakarsanız, “Bu konu üzerine okumaya değecek hiçbir şey yazılmamıştır,” uyarısını görürsünüz.[1] Bu bölümde açıklayacağım üzere, ben daha iyimserim!

Düşünürler binlerce yıl boyunca bilinç gizemi üzerine kafa yormuş olsalar da YZ'lerin yükselişi birden bu konuya –özellikle de hangi zeki varlığın öznel deneyimi olduğunu tahmin etme sorununa- aciliyet kazandırmıştır. Bölüm 3'te de gördüğümüz üzere, zeki makinelerin bir tür hakka sahip olup olmamaları gerekliliği makinelerin bilinçli olup olmadığı ve acı ile mutluluk gibi hisleri yaşayıp yaşayamadığına önemli ölçüde bağlıdır. Bölüm 7'de tartıştığımız gibi, zeki varlıkların deneyimlere sahip olup olmadıklarını bilmeden pozitif deneyimleri maksimize etmeye dayanan faydacı bir etik geliştirmek umutsuzdur. Bölüm 5'te bahsedildiği üzere, bazı insanlar köle-sahip suçluluk duygusundan kaçınmak için robotlarının bilinçsiz olmasını tercih edebilirler. Öte yandan biyolojik sınırlardan kurtulmak için zihinlerini yüklerlerse tam tersini de isteyebilirler: Sonuçta, eğer bilinçsiz bir zombiyse, yani yüklenmiş siz olmak hiçbir şey hissettirmeyecekse, sizin gibi konuşan ve hareket eden bir robota kendinizi yüklemenin anlamı nedir ki? Bu, arkadaşlarınızın öznel deneyiminizin öldüğünü fark etmeyeceği bir şekilde intihar etmeye benzemez mi?

Uzun dönemli kozmik gelecek yaşam için (Bölüm 6), neyin bilinçli neyin bilinçsiz olduğunu anlamak temel odaktır: Eğer teknoloji, zeki yaşamın evrenimizde milyarlarca yıl boyunca gelişmesine olanak sağlayacaksa, yaşamın bilinçli ve ne olup

bittiğini takdir edebilir olduğundan nasıl emin olabiliriz? Eğer olamazsak, ünlü fizikçi Erwin Schrödinger'in sözleriyle, "Boş sandalyelere oynanan, kimse için var olmayan, bu yüzden de aslında hiç var olmayan bir oyun olmaz mıydı bu?"[2] Diğer bir deyişle, eğer yanlışlıkla bilinçli sandığımız yüksek teknoloji bir gelecek nesil oluşturursak, devasa kozmik birikimimizi hiçbir şeye, astronomik bir yer israfına dönüştürdüğümüz hakiki bir zombi kıyameti olmaz mı bu?

Bilinç Nedir?

Bilinç hakkında pek çok sav ışıktan daha çok ısı üretir çünkü taraflar ayrı tellerden çalarlar ve B-kelimesinin farklı tanımlarını kullandıklarının farkında değildirler. "Yaşam" ve "zekâ" konusunda olduğu gibi, "bilinç" kelimesinin de tartışmasız, doğru bir tanımı yoktur. Onun yerine, pek çok tanım birbirleriyle rekabet eder. Bunların arasında, duyguluk, uyanıklık, kendinin farkında olma, duyusal girdiye erişim ve bilgiyle bir anlatı oluşturma yetisi gibi tanımlar vardır.[3] Gelecek zekâsı keşfimizde, şu ana kadar var olagelmiş hiçbir biyolojik bilinçle kısıtlanmamış, üst seviyede geniş ve kapsayıcı bir bakış edinmek istiyoruz. Bu yüzden de Bölüm 1'de verdiğim ve kitap boyunca kullandığım tanım oldukça geniştir.

bilinç = *öznel deneyim*

Diğer bir deyişle, sizin gibi olmak şu an size bir şey hissettiriyorsa, bilinçlisiniz demektir. Bir önceki bölümde YZ kaynaklı tüm soruların özünde işte bu bilinç tanımı yatmaktadır: Prometheus, AlphaGo ya da sürücüsüz Tesla olmak bir şey hissettirir mi?

Bilinç tanımımızın ne kadar geniş olduğunu anlamak için davranış, algı, kendinin farkında olmak, duygular ya da dikkat gibi kelimelerden hiç bahsetmediğime dikkat edin. O yüzden de tanım gereği, rüya gördüğünüzde, tam uyanmamış ve duyusal girdiye erişiminiz olmadığında bile bilinçlisinizdir. Benzer biçimde, acıyı deneyimleyen herhangi bir sistem, hareket edemese bile, bu anlamda bilinçlidir. Bizim tanımımız gelecekteki bir YZ sisteminin de yalnızca bir yazılım olarak var olsa ve hiçbir sensör ya da robot bedene bağlı olmasa bile, bilinçli olabileceği ihtimalini açık bırakmaktadır.

Bu tanımla, bilinç konusunu umursamamak zordur. Yuval Noah Harari'nin kitabı *Homo Deus*'ta dediği gibi:[4] "Eğer herhangi bir bilim insanı öznel deneyimlerin alakasız olduğunu ileri sürerse, işkence ya da tecavüzün hiçbir öznel deneyime referans vermeden neden yanlış olduğunu açıklamak zorunda kalır." Böyle bir referans olmadan, elimizde olan tek şey fizik yasalarına göre temel parçacıkların dolanıp durmasıdır. Bunda yanlış olan ne ki?

Sorun Nedir?

Peki, bilinç hakkında anlamadığımız şey tam olarak nedir? Bu konu üzerine en çok düşünen kişilerden biri de oyunbaz gülümsemesi ve siyah deri ceketi –bu ceketi karım o kadar sevmişti ki bana bir benzerini Noel hediyesi olarak aldı– olmadan göremeyeceğiniz Avustralyalı ünlü filozof David Chalmers'tır. Uluslararası Matematik Olimpiyatları'nda finale kalmış olmasına rağmen –ve üniversitedeki tüm A notlarına rağmen tek B notu felsefeye giriş dersi olsa da– gönlünü felsefeye kaptırmıştı. Gerçekten de bütün sorunlara ve ihtilaflara rağmen azimli gözüküyordu. Cevap verme gereği bile hissetmeden kendi çalışmaları hakkında

bilgisiz ve yanlış eleştirileri kibarca dinleme yeteneği beni her zaman büyülemiştir.

David'in de dediği gibi, zihnin temelde iki ayrı gizemi vardır. İlki beynin bilgiyi nasıl işlediği gizemidir. Bunlara David "kolay" problemler der. Mesela, beyin duyusal girdilerle nasıl ilgilenir, onları nasıl yorumlar ve nasıl cevaplar? Dili kullanarak iç durumu hakkında nasıl haber verir? Bu sorular aslında oldukça zor olsa da bizim tanımımızla bilincin değil, zekânın gizemleridir: Bir beynin nasıl hatırladığını, hesapladığını ve öğrendiğini sorarlar. Dahası, kitabın ilk kısmında YZ araştırmacılarının bu "kolay problemlerin" pek çoğunu makinelerle çözmek konusunda nasıl ciddi ilerleme katettiklerini görmüştük: Go oynamaktan araba sürmeye, imajları analiz etmekten doğal dili işlemeye kadar.

Bir de David'in *zor* problemler adını verdiği, neden öznel bir deneyimimiz olduğuna dair ayrı bir gizem vardır. Araba sürerken renkleri, sesleri, duyguları ve kendin olma hissini deneyimleriz. Tamam da neden herhangi bir şey deneyimliyoruz ki? Sürücüsüz bir araç herhangi bir şey deneyimler mi? Sürücüsüz bir arabaya karşı yarışıyorsanız, hem sensörlerden bilgi iletirsiniz hem onu işlersiniz hem de motor tepkileri çıktı olarak alırsınız. Ama öznel olarak sürmeyi ***deneyimlemek*** mantıksal olarak farklı bir şeydir. Bu seçebileceğiniz bir şey midir ve eğer öyleyse, sebebi nedir?

Bu zor bilinç problemine fizik bakış açısından yaklaşıyorum. Benim perspektifimden, bilinçli bir insan yeniden düzenlenmiş besindir. O zaman bir düzenleme bilinçliyken bir diğeri neden değildir? Dahası, fizik bize besinin aslında belirli bir düzende duran büyük sayıda kuark ve elektron olduğunu söylüyor. O zaman hangi parçacık düzenlemeleri bilinçli ve hangileri bilinçsizdir?*

* Alternatif bir bakış açısı *madde ikiliğidir*; yaşayan varlıkların "anima", "élan vital" ya da "ruh" adı verilen fiziksel olmayan bir materyal içerdikleri için cansız olanlardan farklı olması. Madde ikiliği konusunda bilim insanları arasında destek giderek

Fiziğin getirdiği bu perspektifi sevmemin sebebi, binlerce yıl boyunca çözmeye uğraştığımız bu zor problemi bilimsel yöntemlerle uğraşması kolay, daha odaklı bir hale sokabiliyor olmasıdır. Bir parçacık düzenlemesinin niçin bilinçli hissettiğine dair zor *problem* ile başlamak yerine, bazı parçacık düzenleri bilinçli hissederken bazılarının hissetmediğine dair zor *hakikat* ile başlayalım. Mesela, beyninizi oluşturan parçacıkların şu an bilinçli bir düzende olduğunu ancak derin ve rüyasız bir uykudayken olmadığını biliyorsunuz.

Bu fizik perspektifi Görsel 8.1'de göründüğü üzere, bilinç hakkında üç ayrı zor soruya sebep olur. Öncelikle, parçacık düzeninin hangi özellikleri bir fark yaratır? Özel olarak da hangi fiziksel özellikler bilinçli ve bilinçsiz sistemleri birbirinden ayırır? Eğer buna cevap verebilirsek, hangi YZ sistemlerinin bilinçli olduğunu bulabiliriz. Daha yakın gelecekte, doktorların hangi tepkisiz hastaların bilinçli olduğunu tespit etmelerine de yardımcı olabilir.

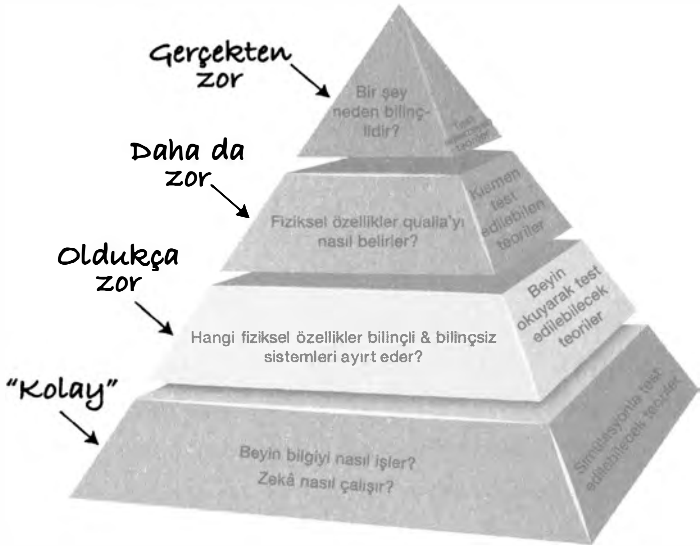
İkincisi, fiziksel özellikler deneyimin ne olduğuna nasıl karar verir? Özel olarak da *qualia*'ya –gülün kırmızılığı, zilin sesi, bifteğin kokusu, mandalinanın tadı ya da iğne batmasının acısı gibi temel bilinç yapı taşlarına- ne karar verir?*

azalmıştır. Bunun sebebini anlamak için, bedeninizin basit fizik yasalarına göre hareket ettiğini söyleyebildiğimiz 10^{29} kuark ve elektrondan yapılmış olduğunu hatırlayalım. Gelecekte bütün parçacıklarınızı takip edebilecek bir teknoloji olacağını varsayalım: Eğer tam olarak fizik yasalarına uyduklarını tespit edersek, sözde ruhunuzun parçacıklarınızın üzerinde hiçbir etkisi yoktur demektir, bu yüzden de bilinçli zihniniz ve hareketlerinizi kontrol etme yetisinin ruhla da hiçbir alakası olmaz. Eğer parçacıklarınız ruhunuz tarafından itildiği için bilinen fizik kurallarına uymuyorlarsa, bu kuvvetlere neden olan yeni varlık, tanımı gereği, geçmişte yeni alanları ve yeni parçacıkları çalıştığımız gibi araştırabileceğimiz fiziksel bir şey olur.

* "*Qualia*" kelimesini sözlükteki anlamına göre kullanıyorum, yani öznel deneyimin tekil anları. Bu da öznel deneyimin kendisi anlamına geliyor, deneyime sebebiyet veren sözde bir madde değil. Bazı insanlar bu kelimeyi farklı anlamlarda kullanıyor.

Üçüncü olarak, bir şey neden bilinçli olur? Diğer bir deyişle, madde kümelerinin neden bilinçli olabildiğine dair derin, keşfedilmemiş bir açıklama var mıdır yoksa bu, dünyanın işleyişiyle ilgili açıklanamayan bir gerçek midir?

Bilgisayar bilimci, MIT'den eski arkadaşım Scott Aaronson, ilk soruya, "oldukça zor problem" (OZP) adını vermişti. Aynı şekilde, Görsel 8.1'de görüldüğü üzere diğer ikisine de "daha da zor problem" (DZP) ve "gerçekten zor problem" (GZP) isimlerini verelim.*



Görsel 8.1: Zihni anlamak bir problemler hiyerarşisini içerir. David Chalmers'ın "kolay" problemler dediği öznel deneyimden bahsetmeden sorulabilir. Tüm sistemlerin değil yalnızca bazılarının bilinçli olduğu gerçeği üç ayrı soruya tekabül eder. Eğer "oldukça zor problemi" tanımlayan soruyu cevaplamak için bir teorimiz varsa, deneysel olarak test edilebilir. Eğer işe yararsa, üstündeki daha zor soruları alt etmek için onun üzerinden gidebiliriz.

* Başta GZP'ye "çok zor problem" demiştim ama bu bölümü David Chalmers'a gösterdiğimde, bana "*gerçekten zor problem*" olarak değiştirme önerisinde bulundu. Böylece şu anlama gelecekti: "İlk iki problem (en azından böyle söyleyelim) bana göre zor problemin gerçekten parçaları değilken, üçüncü problem öyle, bu yüzden de benim kullanımıma uyması için 'çok zor' yerine 'gerçekten zor' kalıbını kullanabilirsin."

Bilinç Bilimin İlerisinde midir?

İnsanlar bana bilinç araştırması umutsuz bir vakit kaybıdır dediklerinde, bana sundukları ana argüman bunun "bilim dışı" olduğu ve hep öyle kalacağıdır. Ama bu gerçekten doğru mu? İtibarlı Avusturyalı-Britanyalı felsefeci Karl Popper şu an geniş ölçüde kabul görmüş, "Eğer yanlışlanabilir değilse bilimsel değildir," veciz sözünü popülerleştirmişti. Diğer bir deyişle bilim, teorileri gözlemlere karşı test etmekten ibarettir: Eğer bir teori ilkesel olarak bile test edilemiyorsa, onu yanlışlamak mantıksal olarak imkânsızdır, bu yüzden de Popper'ın tanımına göre bu bilimsel değildir demektir.

Peki, Görsel 8.1'deki bu üç bilinç sorusunu cevaplayan bir bilimsel teori olabilir mi? Lütfen cevabın kocaman bir EVET! olduğuna sizi ikna etmeme izin verin. En azından oldukça zor problem için: "Hangi fiziksel özellikler bilinçli ve bilinçsiz sistemleri birbirinden ayırır?" Herhangi bir fiziksel sistemi ele aldığımızda birinin sistemin bilinçli olup olmadığına "evet," "hayır" ya da "emin değilim" diye cevap veren bir teorisi olduğunu düşünelim. Beyninizin farklı kısımlarındaki bilgi işlemin bir kısmını ölçen bir cihaza sizi bağlayalım ve elde edilen veriyi, hangi kısımlarının bilinçli olduğunu tahmin etmek için bilinç teorisini kullanan ve sonrasında da Görsel 8.2'deki gibi ekrana gerçek zamanlı olarak tahminini yansıtan bir bilgisayar programına verelim. İlk olarak bir elmayı düşünelim. Ekran beyninizde farkında olduğunuz, elmayla ilgili bir bilgi olduğunu size bildiriyor ancak beyin kökünüzde kalp atışınızla ilgili, farkında olmadığınız bir bilginin daha olduğunu da söylüyor. Etkilenir misiniz? Teorinin ilk iki tahmini doğru olsa bile, biraz daha katı test yapmaya karar veriyorsunuz. Annenizi düşünüyorsunuz ve bilgisayar size beyninizde anneniz hakkında bilgi olduğunu ama bunun farkında olmadığınızı söy-

lüyor. Teori yanlış bir tahminde bulunuyor, bu da teorinin yanlış olduğu ve bilim tarihinin çöplüğünde kendine Aristocu mekanik, ışık veren eter, dünya merkezli kozmoloji ve diğer sayısız hatalı fikirle birlikte yer bulacağı anlamına geliyor. İşte size önemli bir nokta: Teori yanlış olsa bile, *bilimseldir*! Bilimsel olmasaydı, test edemez ve üzerini çizemezdiniz.

Bazıları bu sonucu eleştirebilir ve neyin bilincinde olduğunuz hatta bilinçli olup olmadığınız konusunda hiçbir kanıtları *olmadığını* söyleyebilirler: Bilinçli olduğunuzu söylediğinizi duymuş olsalar da bilinçsiz bir zombi de aynı şeyi muhtemelen söyleyebilirdi. Ancak bu, bilinç teorisini bilimsellik dışı yapmaz çünkü sizinle yer değiştirip *kendi* bilinç deneyimlerini doğru tahmin edip etmediğini test edebilirler.

Öte yandan, eğer teori herhangi bir tahmin yapmayı reddederek, her sorulduğunda "emin değilim" diye cevaplarsa, o zaman test edilemez ve bilimsellik dışıdır. Bu yalnızca bazı durumlara uygulanabilir olduğu, gerekli hesaplamaları pratikte yapmak çok zor olduğu ya da beyin sensörleri iyi olmadığı için gerçekleşebilir. Bugünün en popüler bilimsel teorileri genelde arada bir yerde olurlar, sorularımızın hepsine değil ama bazılarına test edilebilir cevaplar verirler. Mesela, temel fizik teorimiz eş zamanlı olarak hem çok küçük (kuantum mekaniği gerektiren) hem de çok ağır (genel görelilik gerektiren) sistemler hakkında soruları cevaplamayı reddeder çünkü bu durumda hangi matematik denklemleri kullanmamız gerektiğini henüz bilmiyoruz. Bu ana teori mümkün tüm atomların tam kütlelerini tahmin etmeyi de reddeder. Bu durumda, gerekli denklemlere sahip olduğumuzu düşünürüz ancak çözümlerini doğru olarak hesaplamayı başaramayız. Bir teori kendini tehlikeye atarak ve test edilebilir tahminler yaparak ne kadar tehlikeli biçimde yaşarsa o kadar kullanışlıdır ve eğer öldürme çabalarımızın hepsinden sağ çıkarsa biz de onu o kadar

ciddiye alırız. Evet, bilinç teorisinin yalnızca *bazı* tahminlerini test edebiliriz ama bu *tüm* fizik teorileri için böyledir. Bu yüzden de test edemediğimiz hakkında mızmızlanmakla vakit kaybetmeyelim, onun yerine neyi test *edebilirsek* onu test etmeye başlayalım!

Özet olarak, hangi fiziksel sistemin bilinçli olduğunu (oldukça zor problem) tahmin eden herhangi bir teori bilimseldir, yeter ki beynin sonuçlarından hangisinin bilinçli olduğunu tahmin edebilsin. Ancak test edilebilirlik konusu Görsel 8.1'deki üst sorular için daha belirsiz bir hale gelir. Bir teorinin nesnel olarak kırmızı rengini nasıl deneyimlediğinizi tahmin etmesi ne anlama gelir? Bir teori neden bilinç diye bir şey olduğunu açıklama niyetindeyse, bunu deneysel olarak nasıl test edersiniz? Sırf bu sorular zor diye onlardan kaçmak zorunda değiliz ve aşağıda da onlara zaten döneceğiz. Ama birkaç tane alakalı ve cevaplanmamış soruyla karşılaştığımızda, en kolay olanını ilk ele almanın akıllıca olduğunu düşünüyorum. Bu sebeple de MIT'deki bilinç araştırmam tam olarak Görsel 8.1'deki piramidin en altına odaklanıyor. Bu stratejiyi geçtiğimiz günlerde Princeton'dan fizikçi arkadaşım Piet Hut'la tartıştım. O da altını inşa etmeden piramidin en tepesini yapmaya kalkışmanın, kuantum mekaniğinin bize deneylerimizin sonuçlarını tahmin etme şansı tanıyan matematik temeli olan Schrödinger denklemini keşfetmeden kuantum mekaniğinin yorumu hakkında endişelenmeye benzediğini söyleyerek gülmüştü.

Neyin bilimin ötesinde olduğunu tartışırken, cevapların zamana bağlı olduğunu hatırlamak önemlidir! Dört yüzyıl önce, Galileo Galilei matematik temelli fizik teorilerinden öylesine etkilenmişti ki doğayı "matematik dilinde yazılmış bir kitap" olarak tanımlamıştı. Eğer bir üzüm ve bir fındığı fırlatırsa, gidiş yollarını ve ne zaman yere çarpacaklarını da doğru biçimde tahmin edebilirdi. Fakat birinin neden yeşil, diğerinin neden kahverengi ya da birinin neden yumuşak ve diğerinin sert olduğuna dair

bilgisi yoktu – dünyanın bu özellikleri o zamanki bilimin erişebileceğinin ötesindeydi. Ama daima değil! James Clerk Maxwell kendi adını verdiği denklemleri 1861 yılında keşfettiğinde, ışığın ve renklerin matematiksel olarak anlaşılabileceğini açık biçimde ortaya koymuş oldu. 1925'te keşfedilen, yukarıda bahsettiğimiz Schrödinger denklemi maddenin tüm özelliklerini, neyin yumuşak neyin sert olduğu da dâhil olmak üzere, tahmin etmek için kullanılabilir. Teorik ilerleme çok daha fazla bilimsel tahmini mümkün kılarken, teknolojik ilerleme daha fazla deneysel teste olanak sağlamıştır: Şu an teleskoplar, mikroskoplar ve parçacık çarpıştırıcılarıyla üzerinde araştırmalar yaptığımız neredeyse her şey bir zamanlar bilimin ötesindeydi. Diğer bir deyişle, bilimin kapsamı Galileo'nun zamanından beri tüm olguların minnacık bir parçasından büyük bir yüzdesine, atom altı parçacıkları, kara delikleri ve 13.8 milyar yıl önceki kozmik kökenlerimizi dâhil edecek şekilde ciddi oranda genişledi. Bu da şu soruyu ortaya çıkarıyor: Geriye ne kaldı?

Bana göre, bilinç görmezden gelinen aşikâr gerçektir. Bilinçli olduğunuzu bilmekle kalmazsınız, tam bir kesinlikle bildiğiniz *tek şeydir*. Geri kalan her şey, Galileo'nun zamanında René Descartes'ın dediği gibi, çıkarımdır. Teorik ve teknolojik gelişme zamanla bilinci bile bilimin alanına kesin bir şekilde sokacak mı? Bilmiyoruz. Tıpkı Galileo'nun bir gün ışığı ve maddeyi anlayıp anlamayacağımızı bilmediği gibi.* Tek bir şey kesindir: Denemezsek başaramayız! Bu yüzden de ben ve dünyanın dört bir

* Eğer fiziksel gerçekliğimiz, kitabım *Our Mathematical Universe*'te incelediğim gibi, tamamen matematiksel (geniş anlamda konuşursak, bilgi tabanlı) ise gerçekliğin hiçbir hususu –bilinç bile– bilimin kapsamının dışında değildir. Aslında bu bakış açısından bilincin gerçekten zor problemi, matematiksel bir şeyin nasıl fiziksel hissettirdiğini anlama problemiyle tam olarak aynıdır: Eğer matematiksel bir yapının bir kısmı bilinçli ise, kalan kısmını haricî fiziksel dünyası olarak deneyimleyecektir.

yanındaki diğer bilim insanları bilinç teorileri oluşturmak ve onları test etmek için elimizden geleni yapıyoruz.



Görsel 8.2: Bir bilgisayarın beyninizde işlenen bilgiyi ölçtüğünü ve bilinç teorisine göre hangi kısımlarının farkında olduğunuzu tahmin ettiğini düşünün. Tahminlerinin doğru olup olmamasına göre, öznel deneyiminizle eşleştirerek bu teoriyi bilimsel olarak test edebilirsiniz.

Bilinç Hakkında Deneysel İpuçları

Şu esnada kafamızın içinde çok fazla bilgi işlem yapılıyor. Hangileri bilinçli, hangileri bilinçsiz? Bilinç teorilerini ve ne tahmin ettiklerini incelemeden önce, geleneksel düşük teknolojili ya da hiç teknolojik olmayan gözlemlerden son teknoloji beyin ölçümlerine dek bize deneylerin şu ana dek ne öğrettiğine bakalım.

Hangi Davranışlar Bilinçlidir?

Eğer 32 ile 17'yi kafadan çarparsanız, hesaplamanızın pek çok iç çalışmasının bilincindesinizdir. Ancak diyelim ki bunun yerine size Albert Einstein'ın bir portresini gösterdim ve sizden öznenin ismini söylemenizi istedim. Bölüm 2'de gördüğümüz gibi bu da bir hesaplama görevidir. Beyniniz, girdisi gözlerinizden gelen yüksek

sayıda piksel renk ve çıktısı ağzınızı ve ses tellerinizi kontrol eden kaslara giden bilgi olan bir fonksiyonu değerlendirmektedir. Bilgisayar bilimciler bu göreve "konuşma sentezi" ardından gelen "görsel sınıflandırma" adını verirler. Çarpma işleminden çok daha karmaşık olsa da hiç efor sarf etmeden ya da bu işi *nasıl* yaptığınıza dair detayların bilincinde olmadan, bu hesaplamayı çok daha hızlı bir şekilde yapabilirsiniz. Öznel deneyiminiz yalnızca resme bakmayı, tanıma hissi deneyimlemeyi ve kendinizin "Einstein" dediğinizi duymanızı kapsar.

Psikologlar uzun zamandır göz kırpma refleksinden nefes almaya, uzanmaya, bir şeyi tutmaya ve dengenizi sağlamaya kadar bir dizi başka görevleri ve davranışları da bilinçsiz biçimde yapabileceğinizi uzun süredir biliyordu. Tipik olarak, ne yaptığınızın bilincindesinizdir ancak nasıl yaptığınızı bilmezsiniz. Öte yandan, aşina olunmayan durumları, öz kontrolü, karmaşık mantık kurallarını, soyut muhakeme ya da dilin manipülasyonunu içeren davranışlar bilinçli olma eğilimi gösterir. Bunlar *bilincin davranışsal bağıntıları* olarak bilinir ve psikologların "Sistem 2" adını verdiği çaba gerektiren, yavaş ve kontrollü düşünme şekliyle yakından ilişkilidir.[5]

Ayrıca yürümek, yüzmek, bisiklete binmek, araba sürmek, yazmak, tıraş olmak, ayakkabı bağlamak, bilgisayar oyunu oynamak ve piyano çalmak gibi pek çok rutini bilinçliden bilinçsize uzun pratikler sonucunda çevirebildiğiniz de bilinmektedir.[6] Hakikaten, uzmanların kendi uzmanlıklarını en iyi, ne olduğunun yalnızca üst bir seviyede bilincinde oldukları ve işi nasıl yaptıklarına dair alt seviye detaylar hakkında habersiz oldukları "akış" durumundayken yaptıkları da bilinmektedir. Mesela, bir sonraki cümleyi, sanki okumayı yeni öğrenmiş gibi her harfin bilincinde olarak okumayı deneyin. Metni kelimeler ve fikirler

seviyesinde okuduğunuz zamana göre ne kadar yavaş olduğunuzu fark ediyor musunuz?

Gerçekten de bilinçsiz bilgi işlem mümkün olmanın da ötesinde istisnadan ziyade kaide gibidir. Bilimsel kanıtlar duyu organlarımızdan her saniye beynimize giren yaklaşık 10^7 bitlik bilgiden yalnızca 10 ila 50 bitlik ufacık bir kısmının bilincinde olabileceğimizi söylüyor.[7] Bu da sürekli bilincinde olduğumuz bilgi işlemenin yalnızca buzdağının görünen kısmı olduğunu gösteriyor.

Bir arada değerlendirildiğinde, bu ipuçları bazı araştırmacıları bilinçli bilgi işlemin tüm beyinden karmaşık veri analizi gerektiren en önemli kararlarla ilgilenen, zihnimizin CEO'su olarak düşünülebileceğini ileri sürmeye itmiştir.[8] Bu da tıpkı bir şirketin CEO'su gibi neden astlarının uğraştığı her şeyi bilerek dikkatini dağıtmak istemediğini de açıklar ancak eğer isterse de onları bulabilir. Bu seçici dikkati hareket halinde deneyimlemek için, "isterse" kelimesine yeniden bakın: "i" harfinin üzerindeki noktaya bakışınızı sabitleyin, gözlerinizi oynatmadan dikkatinizi noktadan tüm harfe, sonra da tüm kelimeye kaydırın. Retinanızdan gelen veri aynı olsa da bilinçli deneyiminiz değişir. CEO metaforu da uzmanlığın nasıl bilinçsizi oluşturduğunu açıklar: Okumayı ve yazmayı güçlükle öğrendikten sonra CEO, yeni yüksek derece zorluklara odaklanabilmek için bu rutin görevleri bilinçaltı astlarına paslar.

Bilinç Nerededir?

Zekice deneyler ve analizler bilincin yalnızca belirli davranışlarla değil, ayrıca beynin belirli kısımlarıyla da sınırlı olduğunu gösteriyor. Peki, asıl şüpheliler kim? İlk ipuçlarının çoğu beyin lezyonları olan hastalardan geldi: kazalar, inmeler, tümörler ya

da enfeksiyonlardan dolayı ortaya çıkan bölgesel beyin hasarı. Fakat bu genelde netice vermiyordu. Mesela, beynin arkasındaki lezyonların körlüğe yol açtığı gerçeği bu bölgenin görsel bilinç bölgesi olduğu anlamına mı gelir yoksa tıpkı ilk önce gözden geçtiği gibi sonradan bilinçli olacağı yere giderken görsel bilgi buradan mı geçer?

Lezyonlar ve tıbbi müdahaleler bilinçli deneyimin konumunu tam olarak tespit edememiş olsa da seçenekleri azaltmaya yardımcı oldular. Mesela, o an orada gerçekleştiği için acıyı elimde hissediyor olsam da acı deneyiminin başka bir yerde gerçekleşmesi gerektiğini biliyorum çünkü bir cerrah elime hiçbir şey yapmadan el acımı dindirebilmişti: Omzumdaki sinirlerime anestezi yapmıştı. Dahası, uzuvları olmayan bazı kişiler var olmayan ellerinden geliyormuş gibi hissettikleri bir hayalet ağrı deneyimlerler. Başka bir örnek olarak, yalnızca sağ gözümle baktığımda görüş alanımın bir kısmını kaybettiğimi fark ettim; bir doktor retinamın yerinden oynadığına kanaat getirdi ve yerine taktı. Öte yandan, çeşitli beyin lezyonlarına sahip hastalar, görüş alanlarının yarısından bilgi alamadıkları ama bunun da farkında olmadıkları *yarı-ihmal* deneyimlerler. Mesela, tabaklarının sol yanındaki yemeği fark etmeyip yemezler. Sanki dünyalarının yarısına dair bilinçleri kaybolmuştur. Fakat bu hasar görmüş beyin alanları mekânsal deneyim oluşturmakla mı görevliydiler yoksa tıpkı benim retinam gibi bilinç bölgesine mekânsal bilgi mi gönderiyorlardı?

Öncü Amerikalı-Kanadalı beyin cerrahı Wilder Penfield, 1930'larda şu anda *somatosensori korteks* adı verilen belirli beyin bölgelerini elektrikle uyardığında, beyin cerrahisi hastalarının vücutlarının farklı bölgelerine dokunulduğunu söylediğini keşfetmişti (bkz. Görsel 8.3).[9] Ayrıca artık *motor korteks* adı verilen bölgeleri uyardığında vücutlarının farklı kısımlarını istemsiz biçimde hareket ettiklerini de bulmuştu. Ama bu, beynin bu

alanlardaki bilgi işleminin dokunma ve hareket bilincine denk düştüğü anlamına mı gelir?

Neyse ki modern teknoloji bize artık çok daha detaylı ipuçları veriyor. Yaklaşık yüz milyar nöronumuzun her birinin her bir ateşlemesini ölçmenin yakınında bile olmasak da beyin okuma teknolojisi, fMRI, EEG, MEG, ECoG, ePhys ve floresan voltaj duyumlama gibi göz korkutucu isimlere sahip teknolojilerle birlikte hızla ilerliyor. Fonksiyonel manyetik rezonans görüntüleme anlamına gelen fMRI yaklaşık her saniye milimetre çözünürlüğünde beyninizin 3D haritasını yapmak için hidrojen çekirdeğinin manyetik özelliklerini ölçer. EEG (elektroensefalografi) ve MEG (manyetoensefalografi) beyninizin saniyede bin kez haritasını çıkarmak için başınızın dışındaki elektrik ve manyetik alanı ölçer ama birkaç santimetreden daha kısa şeyleri ayırt etmede başarısızdır ve düşük çözünürlüklüdür. Eğer biraz hassas biriyseniz, bu üç tekniğin de beynin içine müdahale etmediğini bilmek sizi rahatlatacaktır. Eğer kafatasınızın açılmasından rahatsız olmazsanız başka seçenekleriniz de var. ECoG (elektrokortikografi) beyninizin yüzeyine yaklaşık yüz kablo takılmasını içerirken ePhys (elektrofizyoloji), bine kadar eş zamanlı konumdan voltaj kaydetmek için beyninizin derinliklerine bazen insan saçından bile daha ince olan mikro kablolar yerleştirir. Çoğu epilepsi hastası için beynin hangi kısmının nöbetlere neden olduğunu ve kesip çıkarılması gerektiğini bulmakta ECoG kullanılırken ve sinirbilimcilerin bu esnada onlar üzerinde bilinç deneyleri yapmasına nezaketle izin verirken hastanede günlerini harcarlar. Son olarak, floresan voltaj duyumlama da nöronların ateşlenirken ışık yaymasını sağlayacak şekilde genetik biçimde manipüle edilmesini içerir. Böylece aktiviteleri mikroskopla ölçülebilir. Tüm bu tekniklerin dışında, en azından şeffaf beyinleri olan hayvanlarda, en yüksek oranda

nöronu hızla takip etme potansiyeli vardır: üç yüz iki nöronuyla *C. elegans* solucanı ve yüz bin nöronuyla larval zebra balığı gibi.

Francis Crick, Christof Koch'u bilinç üzerine çalıştığı için uyarmış olsa da Christof vazgeçmeyi reddetti ve sonunda Francis'i ikna etmeyi başardı. 1990 yılında, "bilincin sinirsel bağıntıları" (NCC) adını verdikleri şey üzerine hangi beyin süreçlerinin bilinçli deneyime tekabül ettiğini araştıran ufuk açıcı bir makale yazdılar. Binlerce yıl boyunca, düşünürler beyinlerinde işlenen bilgiye yalnızca öznel deneyimleri ve davranışları yoluyla erişebiliyorlardı. Crick ve Koch, beyin okuma teknolojisinin aniden bu bilgiye bağımsız erişim sağladığını ortaya koyarak hangi bilinçli deneyime hangi bilgi işlemenin denk düştüğüne dair bilimsel çalışmaya olanak sağladılar. Gerçekten de teknoloji destekli ölçümler şu anda NCC arayışını sinirbilimin ana akım kısımlarından biri haline getirdi. Bu alandaki binlerce yayın, en prestijli dergilerde bile yer alıyor.[10]

Şimdiye dek varılan sonuçlar neler? NCC dedektifliğinin ne olduğunu anlayabilmek için, önce retinanızın bilinçli mi yoksa görsel bilgiyi kaydeden, işleyen ve öznel görsel deneyiminizin gerçekleştiği beyninizin bir yerindeki bir sisteme gönderen bir zombi sistem mi olduğunu soralım. Görsel 8.4'ün sol panelinde, hangi kare daha koyudur? A mı yoksa B mi? A, değil mi? Hayır, aslında aynı renge sahiptirler, bunu da parmaklarınızın arasındaki küçük deliklerden bakarak görebilirsiniz. Bu da görsel deneyiminizin yalnızca retinanızda olamayacağını kanıtlar. Eğer öyle olsaydı, aynı görünürlerdi.

Şimdi de Görsel 8.4'ün sağ paneline bakın. İki kadın mı görüyorsunuz yoksa bir vazo mu? Eğer yeterince uzun bakarsanız, retinanıza ulaşan bilgi aynı olsa da ikisini de sırayla öznel olarak deneyimlersiniz. Bu iki durum esnasında ne olduğunu ölçerek,

farkın ne olduğunu ayırt edebilirsiniz ve bu fark retinadan kaynaklanmaz çünkü iki durumda da aynı şekilde davranır.

Bilinçli retina hipotezine öldürücü darbe Christof Koch, Stanislas Dehaene ve diğerlerinin öncülüğünde ortaya çıkan "sürekli flaş baskılama" adı verilen bir teknikten gelir: Eğer gözlerinizden birine hızla değişen örüntülerden oluşan karmaşık bir sekansı takip ettirirseniz, diğer gözünüze gösterilen sabit görüntünün hiç farkında olmayacağı şekilde görsel sisteminizin dikkatini dağıtabilirsiniz.[11] Özet olarak, retinanızda görsel bir imaja sahip olsanız da onu deneyimlemeyebilirsiniz ve (rüya görürken) retinanızda olmasa bile o imajı deneyimleyebilirsiniz. Bu da, yüz milyonlarca nöronu içeren karmaşık hesaplamaları gerçekleştirseler bile, iki retinanızın bir video kameradan daha çok görsel bilince sahip olmadığını kanıtlar.

NCC araştırmacıları, dengesiz görsel/duyusal illüzyonları ve diğer numaraları, sürekli flaş baskılamayı beyin bölgelerinizden hangisinin bilinçli deneyimlerinizden sorumlu *olduğunu* tespit etmekte kullanırlar. Temel strateji, bilinçli deneyiminiz hariç, neredeyse her şeyin (duyusal girdi de dâhil olmak üzere) aynı olduğu iki durumda nöronlarınızın ne yaptığını kıyaslamaktır. Beyninizin farklı davrandığı ölçülmüş kısımları NCC olarak tanımlanır.

Böylesi NCC araştırmaları, besinleri optimal olarak nasıl sindireceğimizi hesaplayan, yarım milyar nöronla bağırsak kökenli sinir sisteminizin konumu olsa da bilincinizin hiçbir kısmının karnınızda olmadığını da kanıtlamıştır; açlık ve mide bulantısı gibi hisler beyninizde üretilir. Benzer biçimde, bilincinizin hiçbir kısmı beyin kökünüzde, omuriliğinize bağlanan ve nefes alma, kalp ritmi ve tansiyonu kontrol eden beyninizin alt kısmında, yer almamaktadır. Daha da ilginç biçimde, bilinciniz, nöronlarınızın üçte ikisine sahip olan beyinciğinize (Görsel 8.3) de uzanmaz:

beyincikleri yok olan hastalar sarhoş gibi geveleyerek konuşurlar ve yalpalayarak hareket ederler, ama tamamen bilinçlidirler.

Beyninizin hangi kısımlarının bilinçten sorumlu *olduğu* sorusu halen tam olarak yanıtlanmamıştır ve tartışmalıdır. Bazı yeni NCC araştırmaları, bilincinizin talamus (beyninizin orta kısmına yakın) ve korteksinizin arka kısmını (düzleştirildiğinde büyük bir peçetenin alanına sahip olacak altı katlı buruşuk bir yüzeyden oluşan beynin dış katmanı) içeren "sıcak bölgede" yer aldığını ileri sürer.[12] Aynı araştırma, tartışmalı biçimde, başın en arkasındaki birincil görsel korteksin bir istisna olduğunu, gözleriniz ve retinalarınız kadar bilinçsiz olduğunu da öne sürmektedir.

Bilinç Ne Zamandır?

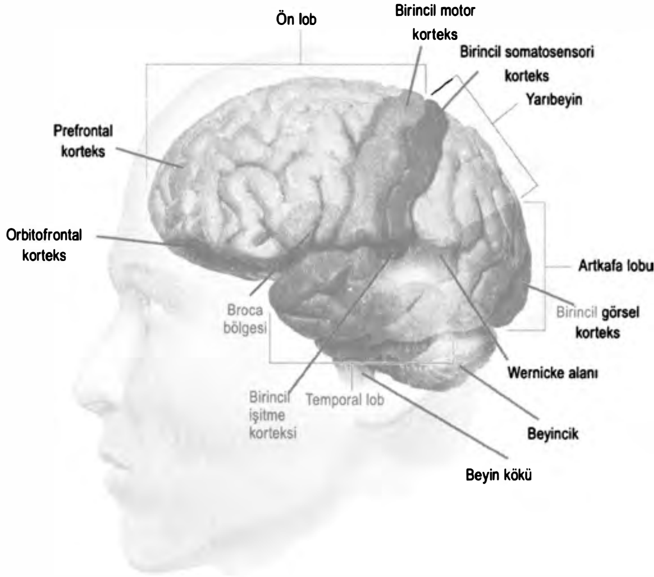
Şu ana dek ne tür bilgi işlemin bilinçli olduğuna ve bilincin nerede gerçekleştiğine bağlı olarak deneysel ipuçlarını değerlendirdik. Peki, bu *ne zaman* gerçekleşir? Ben çocukken, olaylar gerçekleşirken, hiçbir gecikme olmaksızın, o olayların bilincinde oluruz diye düşünürdüm. Öznel olarak bana hâlâ öyle geliyor olsa da bu kesinlikle doğru olamaz çünkü beynimin duyu organlarımdan gelen bilgiyi işlemesi için zaman geçmesi gerekiyor. NCC araştırmacıları bu sürenin ne kadar olduğunu dikkatle ölçtüler ve Christof Koch'un vardığı sonuca göre, karmaşık bir nesneden gelen ışığın gözünüze girmesinden siz onu olduğu şey olarak bilinçli bir şekilde algılayana dek bir saniyenin çeyreği kadar zaman geçmektedir.[13] Bu da saatte elli mil ile bir otoyolda aracınızı sürerken birkaç metre önünüzde bir sincap görürseniz, bir şey yapmak için artık çok geç olduğu anlamına gelir çünkü çoktan onu ezmişsinizdir!

Özet olarak, bilinciniz geçmişte yaşar. Christof Koch'un tahminine göre de dış dünyayı çeyrek saniye geriden takip eder.

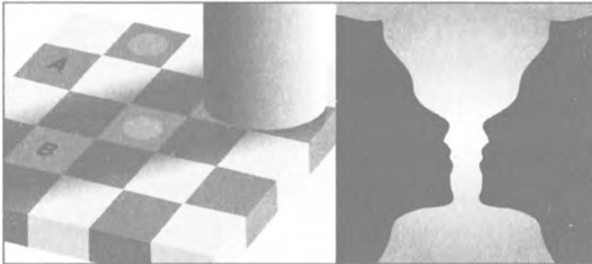
Merak uyandırıcı bir biçimde, bazı şeylere onların bilincinde olamayacağınız kadar kısa sürede tepki gösterirsiniz. Bu da en hızlı tepkilerimizden sorumlu bilgi işlemenin bilinçsiz olması gerektiğini kanıtlar. Mesela, gözünüze yabancı bir nesne yaklaşırsa, göz kırpma refleksiniz göz kapağınızı saniyenin onda biri içinde kapar. Sanki beyin sistemlerinizden biri uğursuz bir bilgiyi görsel sisteminizden almış, gözünüzün çarpma riski altında olduğunu hesaplamış, göz kaslarınıza kırpma mesajını yollamış ve aynı anda da beyninizin bilinçli kısmına, "Hey, gözlerimizi kırpacağız," diye başka bir mesaj atmış gibidir. Bu mesaj okunup da bilinçli deneyiminize eklendiğinde, kırpma çoktan gerçekleşmiş olacaktır.

Gerçekten de bu mesajı okuyan sistem devamlı olarak tüm bedeninizden mesajlarla bombalanmaktadır, bazıları diğerlerine göre daha geç gelir. Sinir sinyallerinin parmaklarınızdan beyninize ulaşması yüzünüze ulaşmasından daha uzun sürer çünkü mesafe daha uzundur. Ayrıca görselleri analiz etmeniz sesleri analiz etmenizden daha uzun sürer çünkü görseller daha karmaşıktır. Bu yüzden de Olimpiyat yarışları görsel bir işaretle değil silah sesiyle başlar. Fakat burnunuza dokunursanız, bilinçli olarak burnunuzda ve parmağınızda eş zamanlı bir his deneyimlersiniz ve eğer ellerinizi çırparsanız, bu çırpmayı aynı anda görür, duyar ve hissedersiniz.[14] Bu da bir olay hakkındaki bütün bilinçli deneyiminizin, o son hantal mesaj içeri girip analiz edilmeden önce yaratılmadığı anlamına gelir.

Fizyolog Benjamin Libet'in başını çektiği ünlü bir NCC deneyleri ailesi bilinçsiz gerçekleştirebileceğiniz bu tür aksiyonların göz kırpma, masa tenisi oynama gibi hızlı cevaplarla sınırlı olmadığını, özgür iradeye atfedebileceğiniz çeşitli kararları da içerdiğini göstermiştir. Beyin ölçümleri, bazen kararınızı siz onu verdiğinizin bilincine ermeden önce tahmin edebilir.[15]



Görsel 8.3: Görsel, işitsel, somatosensori ve motor korteksleri sırasıyla görme, duyma, dokunma ve hareket aktivasyonuyla ilgilidir ancak bu, görme, duyma, dokunma ve hareketin *bilincinin* oluştuğu yer oldukları anlamına gelmez. Zaten son araştırmalar da birincil görsel korteksin, beyincik ve beyin köküyle birlikte, tamamen bilinçsiz olduğunu ortaya koymuştur. Görsel Lachina'dan alınmıştır. (www.lachina.com).



Görsel 8.4: Hangi kare daha koyudur: A mı B mi? Sağda gördüğünüz nedir: bir vazo, iki kadın, yoksa sırasıyla ikisi de mi? Bunun gibi illüzyonlar görsel bilincinizin gözlerinizde ya da görsel sisteminizin erken aşamalarında olamayacağını gösterir çünkü yalnızca resimde olan şeye bağlı değildir.

Bilinç Teorileri

Henüz bilinçliliği tam olarak anlamamış olsak da çeşitli veçhelerine dair inanılmaz miktarda bilgiye sahip olduğumuzu gördük. Tüm bu veri *beyinlerden* geliyor, peki bize *makinelerdeki* bilinç hakkında bir şeyler öğretebilir mi? Bu mevcut deneysel alanımızın çok ötesinde, bilgiye dayalı bir tahmin gerektirir. Diğer bir deyişle, bir *teoriye* ihtiyacımız vardır.

Neden bir Teori?

Neden olduğunu anlamak için, bilinç teorilerini yer çekimi teorileriyle kıyaslayalım. Bilim insanları Newton'ın yer çekimi teorisini ciddiye almaya başladılar çünkü verdiklerinden daha fazlasını alıyorlardı: Bir peçeteye sığabilecek kadar basit denklemler, her yer çekimi deneyinin sonucunu doğru biçimde tahmin edebiliyordu. Bu yüzden de test edildiği alanın çok ötesinde de tahminlerini ciddiye almaya başlamışlardı ve bu cesur tahminler milyonlarca ışık yılı ötedeki kümelerde bulunan galaksilerin hareketlerini bile açıklıyordu. Ancak tahminler Merkür'ün Güneş etrafındaki hareketi için ufak bir miktar sapmıştı. Bilim insanları da o noktada Einstein'ın iyileştirilmiş yer çekimi teorisini, genel göreliliği ciddiye almaya başladılar çünkü çok daha incelikli ve ekonomikti ve Newton'un teorisinin yanlış tahmin ettiği şeyleri bile doğru biliyordu. Sonuç olarak kara delikler, uzayzamanın dokusundaki yer çekimi dalgaları gibi egzotik olgular ve sıcak, ateşli başlangıcından itibaren evrenimizin genişlemesi gibi test edildiği alanın ötesinde de tahminlerini ciddiye almaya başlamışlardı. Bunların tümü sonradan deneylerle teyit edilmişti.

Benzer biçimde, denklemleri peçeteye sığacak kadar sade olabilen bilincin matematiksel bir teorisi, beyin üzerinde uyguladığımız tüm deneylerin çıktılarını başarılı bir şekilde tahmin

edebilirse, o zaman yalnızca teoriyi değil beynin ötesinde bilinç için yaptığı tahminleri de ciddiye almaya başlayabiliriz; mesela makinelerdekini.

Fizik Perspektifinden Bilinç

Bazı bilinç teorileri antik çağlara kadar uzansa bile çoğu modern teori nöropsikoloji ve sinirbilime dayanır ve bilinci beyinde gerçekleşen sinirsel olaylar olarak açıklamaya ve tahmin etmeye çalışır.[16] Bu teoriler bilincin sinirsel bağıntıları için başarılı tahminlerde bulunmuş olsalar da makine bilinci için tahminler yapmaya aday değildirler. Beyinden makineye bu atılımı gerçekleştirmek için, NCC'den PCC'ye genelleme yapmamız gerekir: Yani bilinçli hareket eden parçacık örüntüleri olarak tanımlanan *bilincin fiziksel bağıntılarına*. Çünkü bir teori, eğer neyin bilinçli neyin bilinçsiz olduğunu yalnızca temel parçacıklar, kuvvet alanları gibi fiziksel yapı taşlarına referans vererek doğru biçimde tahmin edebiliyorsa, yalnızca beyinler değil, geleceğin YZ sistemleri de dâhil olmak üzere tüm diğer madde düzenleri için de tahminlerde bulunabilir. O zaman fizik perspektifinden bakalım: Hangi parçacık düzenleri bilinçlidir?

Fakat bu gerçekten de başka bir soruyu akla getiriyor: Bilinç kadar karmaşık bir şey nasıl parçacık kadar basit bir şeyden yapılmış olabilir? Bence bunun sebebi parçacıklarının ötesinde ve üzerinde özelliklere sahip olan bir olgu olmasıdır. Fizikte böyle olgulara, "beliren" deriz.[17] Bunu, bilinçten daha basit bir "beliren" olguya bakarak anlamaya çalışalım: ıslaklık.

Bir su damlası ıslaktır ancak aynı su moleküllerinden yapılmış olmalarına rağmen bir buz kristali ve bir buhar bulutu değildir. Neden? Çünkü ıslaklığın özelliği moleküllerin düzenine bağlıdır. Tek bir su molekülünün ıslak olduğunu söylemek hiçbir anlam ifade etmez çünkü ıslaklık olgusu yalnızca pek çok molekülün

olduğu, sıvı dediğimiz bir örüntü halinde dizildikleri durumda ortaya çıkar. Bu yüzden de katılar, sıvılar ve gazların her biri beliren olgudur: Parçalarının toplamından daha fazlasıdırlar çünkü parçacıklarının özelliklerinin ötesinde ve üzerinde özelliklere sahiptirler. Parçacıklarının sahip olmadığı özellikleri vardır.

Tıpkı katılar, sıvılar ve gazlar gibi bilincin de beliren bir olgu olduğunu, parçacıklarının ötesinde ve üzerinde özelliklere sahip olduğunu düşünüyorum. Mesela, derin uykuya girmek, yalnızca parçacıklarını yeniden düzenleyerek bilinci bastırır. Aynı şekilde, eğer donarak ölürsem bu, parçacıklarımı daha talihsiz bir biçimde yeniden düzenleyeceği için bilincim de ortadan kalkar.

Pek çok parçacığı bir araya koyarak sudan beyne kadar herhangi bir şey yaptığınızda, gözlemlenebilir özellikleriyle yeni bir olgu belirir. Biz fizikçiler bu beliren özellikleri araştırmaya bayılırız. Bu özellikler genelde ölçebileceğimiz küçük bir dizi sayıyla tanımlanır; bunlar bir maddenin akışmaz olma durumu, ne kadar sıkıştırılabileceği gibi miktarlardır. Mesela, eğer bir madde katı olacak kadar akışmaz ise ona katı, diğer durumda da akıcı deriz. Eğer akıcı bir madde sıkıştırılamaz ise ona sıvı der, diğer türlü de elektriği ne kadar iyi ilettiğine bağlı olarak gaz ya da plazma diye adlandırırız.

Bilgi Olarak Bilinç

Peki, bilinci ölçecek benzer nicelikler olabilir mi? İtalyan sinirbilimci Giulio Tononi böyle bir Yunan harfi Φ (*Phi*) ile gösterdiği, bir sistemin farklı kısımlarının birbiri hakkında ne kadar çok şey bildiğini ölçen ve "*entegre bilgi*" adını verdiği bir nicelik teklif etmiştir (bkz. Görsel 8.5).

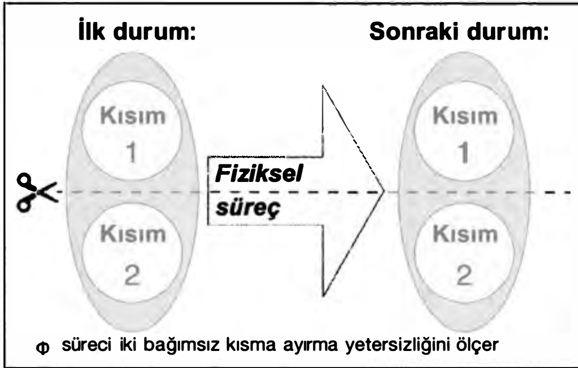
İlk kez Giulio'yla 2014'te Porto Riko'da düzenlediğimiz ve onu Christof Koch'la birlikte davet ettiğim fizik konferansında

tanışmıştım. İlk görüşte beni, Galileo ile Leonardo da Vinci'nin arasında sırıtmayacak bir Rönesans adamı olarak etkilemişti. Sessiz tutumu sanat, edebiyat ve felsefe konusundaki inanılmaz bilgisini saklayamıyordu ve mutfak alanındaki namı kendisinden önde gidiyordu: Kozmopolit bir TV muhabiri bana Giulio'nun birkaç dakika içerisinde nasıl hayatında yediği en lezzetli salatayı yaptığını anlatmıştı. Ağırbaşlı tavrının ardında tüm yerleşik fikir ve tabulara rağmen kanıtın onu götürdüğü yere gitmekten asla korkmayacak bir zihni olduğunu kısa sürede fark ettim. Dünya merkezli evren teorisine meydan okumaması için müesses nizamın yaptığı baskılara rağmen kendi matematiksel hareket teorisinin peşinden giden Galileo gibi Giulio da bugüne kadar yapılmış, matematiksel olarak en kesin bilinç teorisini geliştirmiştir: *entegre bilgi teorisi* (IIT).

On yıllardır bilincin, çeşitli karmaşık şekillerde işlendiği zaman bilginin hissettiği şey olduğunu ileri sürüyorum.[18] IIT de buna katılıyor ve benim belirsiz "çeşitli karmaşık şekiller" cümleciğimi net bir tanımla değiştiriyor: Bilgi işlemenin entegre edilmesi gerekir, yani Φ büyük olmalıdır. Giulio'nun bu konudaki argümanı da basit olduğu kadar güçlüdür: Bilinçli sistemin birleşmiş bir bütüne entegre edilmesi gerekir çünkü iki bağımsız kısım olsaydı, bir değil iki ayrı bilinçli varlık olarak hissederlerdi. Diğer bir deyişle, eğer bir beynin ya da bilgisayarın bilinçli bir kısmı geri kalanıyla iletişim kuramıyorsa, kalan kısım öznel deneyiminin kısmı olamaz.

Giulio ve çalışma ortakları, beynin manyetik uyarılara cevabını ölçmek için EEG kullanarak Φ'nin basitleştirilmiş bir versiyonunu ölçtüler. Yaptıkları "bilinç dedektörleri" oldukça iyi çalışıyordu: Hastaların uyanıkken ya da rüya görürken bilinçli olduklarını ama anestezi altında ya da derin uykudayken bilinçsiz oldukla-

rını tespit etmişti. Hatta *locked-in* sendromuna yakalanmış, hiçbir şekilde hareket edemeyen ya da iletişim kuramayan iki hastanın bilincini keşfetmişti.[19] Bu yüzden de çeşitli hastaların bilinçli ya da bilinçsiz olduğunu keşfedebilecek, gelecekte doktorlar için ümit vadeden bir teknoloji olarak ortaya çıkıyor.



Görsel 8.5: Zamanla bir sistemin ilk durumundan yeni bir duruma dönüşen fiziksel bir süreci ele aldığımızda, *entegre bilgi* süreci bağımsız kısımlara ayırma yetersizliğini ölçmüştür. Eğer her kısmın gelecek durumu yalnızca kendi geçmişine bağlıysa ve diğer kısmın yaptığından bağımsızsa, Φ=0 olur. Tek bir sistem dediğimiz şey aslında birbiriyle hiçbir şekilde iletişim kurmayan iki bağımsız sistemdir.

Bilinci Fiziksel Temele Oturtma

IIT yalnızca, bilgisayar belleğindeki bitler ya da kapalı ve açık durumları olan aşırı basitleştirilmiş nöronlar gibi sonlu sayıda durumu olan ayrık sistemler için tanımlanmıştır. Bu da maalesef IIT'nin sürekli değişen, en geleneksel fizik sistemleri için tanımlı olmadığı anlamına gelir; mesela, bir parçacığın konumu ya da manyetik alanın gücü sonsuz sayıda değer alabilir.[20] Eğer IIT formülünü böyle sistemlere uygulamayı denerseniz, Φ'nin sonsuz olduğu işe yaramaz bir sonuç alırsınız. Kuantum mekanik sis-

temleri ayrık olabilir ancak orijinal IIT, kuantum sistemleri için tanımlanmamıştır. Peki, IIT ve diğer bilgi bazlı bilinç teorilerini nasıl sağlam bir fiziksel temele oturtacağız?

Madde kümelerinin bilgiyle alakalı beliren özelliklere nasıl sahip olabildiğine dair Bölüm 2'de öğrendiğimiz şeylerin üzerine inşa ederek bunu yapabiliriz. Bir şeyin bilgi depolayan bellek cihazı olarak kullanılabilmesi için pek çok uzun ömürlü durumunun olması gerektiğini görmüştük. Ayrıca, *computronium*, hesaplama yapabilen bir madde olmak da karmaşık dinamiklere gereksinim duyar: fizik kuralları rastgele bilgi işlemi gerçekleştirebilecek kadar karmaşık olan yollarda değişmesini sağlamak zorundadır. Son olarak, örneğin bir sinirsel ağın öğrenmek için nasıl güçlü bir madde olabileceğini çünkü yalnızca fizik kurallarına uyarak kendini arzu edilen hesaplamaları yapmakta giderek daha da iyi yapacak şekilde yeniden düzenleyebileceğini görmüştük. Şimdi bir soru daha soruyoruz: Bir madde kümesinin öznel deneyime sahip olabilmesini ne sağlar? Yani hangi koşullar altında bir madde kümesi şu aşağıdaki dört şeyi yapabilir hale gelir?

1. Hatırlama
2. Hesaplama
3. Öğrenme
4. Deneyimleme

İlk üçünü Bölüm 2'de incelemiştik ve şimdi de dördüncünün peşindeyiz. Tıpkı Margolus ve Toffoli'nin rastgele hesaplamalar yapabilen maddeye *computronium* adını vermesi gibi, öznel deneyime sahip olan (duygulu olan) en genel madde için *sentronium* terimini kullanmayı seviyorum.*

* Önceden aynı anlamda "perceptronium" terimini kullanmış olsam da bu isim çok dar bir tanım öne sürüyor çünkü algılar yalnızca duyusal girdilerimize dayalı olarak algıladığımız öznel deneyimlerdir, rüyaları ve içimizde geliştirdiğimiz düşünceleri kapsamaz.

Fakat eğer fiziksel bir olgu ise, bilinç nasıl hiç de fiziksel değilmiş gibi hissedilebilir? Nasıl olur da fiziksel özünden bu kadar bağımsızmış gibi hissedilebilir? Bunun sebebinin aslında içinde bir örüntü olduğu fiziksel özünden bağımsız *olması* olduğunu düşünüyorum! İçinde bulundukları fiziksel maddeden bağımsız örüntülerin pek çok güzel örneğini Bölüm 2'de görmüştük. Mesela dalgalar, anılar ve hesaplar gibi... Parçalarından daha fazla olmakla kalmayıp (beliren), kısımlarından bağımsız, hatta kendi başlarına bir yaşamları olan şeylerdi bunlar. Mesela, geleceğin simüle edilmiş bir zihni ya da bilgisayar oyunu karakterinin Windows, MacOS, Android telefon ya da başka bir işletim sisteminde çalıştığını bilmesinin yolunun olmadığını çünkü maddeden bağımsız olduğunu görmüştük. Bilgisayarının mantık kapılarının transistörlerden, optik devrelerden ya da diğer tür bir donanımdan yapıldığını da söyleyemez. Ya da fiziğin temel yasalarının ne olduğunu – evrensel bilgisayarların inşasına izin verdiği sürece her şey olabilirler.

Özet olarak, bilincin fiziksellik dışı hissettiren fiziksel bir olgu olduğunu çünkü dalgalar ve hesaplamalar gibi olduğunu düşünüyorum: Kendi fiziksel maddesinden bağımsız özelliklere sahiptir. Bu, bilincin bilgi olduğu fikrinin mantıksal devamıdır. Bu da benim çok sevdiğim radikal bir fikre çıkar: Eğer bilinç bilginin çeşitli şekillerde işlendiğinde hissediş şekli ise, maddeden bağımsız olmalıdır; önemli *olan* tek şey bilgi işlemenin yapısıdır, bilgi işlemeyi gerçekleştiren maddenin yapısı değil. Diğer bir deyişle, bilinç, iki kere maddeden bağımsızdır!

Gördüğümüz gibi, fizik uzayzamanda hareket eden parçacıklara denk düşen örüntüleri tanımlar. Eğer parçacık düzenlemesi belirli ilkelere uyarsa, parçacık özünden bağımsız olarak beliren olguya yol açar ve yarattıkları his de tamamen farklıdır. Bunun çok iyi bir örneği *computronium*'da bilgi işlemedir. Ama şimdi biz

bu fikri başka bir seviyeye çıkarıyoruz: *Eğer bilgi işlemenin kendisi belirli ilkelere uyarsa, bilinç dediğimiz yüksek seviye beliren olguya yol açabilir.* Bu bilinçli deneyiminizi maddeden bir değil iki seviye yukarı yerleştirir. Zihninizin fiziksel değilmiş gibi hissetmesine şaşmamalı!

Bu da bir soruya neden olur: Bilgi işlemenin bilinçli olması için uyması gereken bu ilkeler nelerdir? Bilinçliliğin kesin olarak var olması için hangi koşulların *yeterli* olacağını biliyormuş gibi davranmayacağım ama üzerine iddiaya girebileceğim ve araştırmamda incelediğim dört *gerekli* koşul şunlardır:

İlke	**Tanım**
Bilgi ilkesi	Bilinçli bir sistemin azımsanamayacak bir bilgi depolama kapasitesi vardır.
Dinamik ilkesi	Bilinçli bir sistemin azımsanamayacak bir bilgi işleme kapasitesi vardır.
Bağımsızlık ilkesi	Bilinçli bir sistem dünyanın geri kalanından azımsanamayacak ölçüde bağımsızdır.
Entegrasyon ilkesi	Bilinçli bir sistem neredeyse bağımsız kısımlardan oluşamaz.

Dediğim gibi bilincin çeşitli şekillerde işlendiğinde bilginin hissedişi olduğunu düşünüyorum. Bu da bilinçli olmak için bir sistemin bilgiyi depolayıp işleyebilmesi gerektiği anlamına geliyor, yani ilk iki ilkeyi karşılamalı. Belleğin uzun süreli olması da gerekmez: Hafızası bir dakikadan kısa olan ama mükemmel biçimde bilinçli olan Clive Wearing'in şu dokunaklı videosunu izlemenizi öneririm.[21] Bilinçli bir sistemin dünyanın geri kalanından ciddi oranda bağımsız olması gerektiğini de düşünüyorum çünkü diğer türlü bağımsız bir varlığı olduğunu öznel olarak hissedemez. Son olarak da Giulio Tononi gibi, bilinçli sistemin birleşik bir bütün olarak entegre olması gerektiğini düşünüyorum

çünkü iki bağımsız kısımdan oluşursa, bir değil iki ayrı bilinçli varlık olarak hissederlerdi. İlk üç ilke *özerkliğe* işaret eder: Sistemin dışarıdan pek müdahale olmadan bilgiyi koruyup işleyebilme, kendi geleceğine karar verebilme yetisi. Dört ilke hep beraber sistemin özerk olduğu anlamına gelir ama parçaları özerk değildir.

Eğer bu dört ilke doğruysa, yapmamız gereken iş bellidir demektir: Bu ilkeleri kapsayacak matematiksel olarak sağlam teorileri aramalı ve deneysel olarak test etmeliyiz. Ayrıca ek ilkelerin gerekip gerekmediğini tespit etmeliyiz. IIT'nin doğru olup olmadığından bağımsız olarak, araştırmacıların rakip teoriler geliştirmeyi denemeleri ve mevcut tüm teorileri daha iyi deneylerle test etmeleri gerekir.

Bilincin İhtilafları

Bilinç araştırmasının bilimsellik dışı saçmalık ve amaçsız zaman kaybı olup olmadığına dair bitmek tükenmek bilmeyen ihtilafları zaten tartışmıştık. Buna ek olarak, en son bilinç araştırmalarında yeni ihtilaflar da mevcut. Şimdi benim en aydınlatıcı olduğunu düşündüğüm bir tanesini inceleyelim.

Giulio Tononi'nin IIT'si son zamanlarda yalnızca övgü değil eleştiri oklarına da hedef oldu, bu eleştirilerin bazıları da çok sertti. Geçtiğimiz günlerde Scott Aaronson blogunda şunu demişti: "Bana göre, Entegre Bilgi Teorisi'nin yanlış olması –temeline kadar giden sebeplerden ötürü, açıkça yanlış– teoriyi, şu ana kadar ortaya atılmış tüm matematiksel bilinç teorilerinin en üstteki %2'si arasına sokuyor. Bana öyle geliyor ki rakip bilinç teorilerinin neredeyse hepsi o kadar belirsiz, kof ve uysallar ki ancak yanlış olmayı arzu edebilir."[22] Hem Scott hem de Giulio'ya övgü olarak şunu söylemeliyim ki geçenlerde New York Üniversitesi'nde düzenlenen bir atölyede IIT üzerinden tartışmalarını izledim,

saç başa girmediler, aksine birbirlerinin argümanlarını kibarca dinlediler. Aaronson belirli basit mantık kapısı ağlarının aşırı derece yüksek entegre bilgiye (Φ) sahip olduğunu gösterdi ve bilinçli olmadıkları açık olduğu için IIT'nin yanlış olduğunu ileri sürdü. Giulio eğer inşa edilmiş olsalardı, bilinçli *olacaklarını* söyleyerek buna karşı çıktı ve Scott'ın tersi varsayımının, tıpkı bir mezbaha sahibinin hayvanların bilinçli olamayacağını çünkü konuşamadıklarını ve insanlardan çok farklı olduklarını söyleyerek yaptığı biçimde insan merkezci önyargıya sahip olduğunu iddia etti. İkisinin de onayladığı benim analizim, entegrasyonun bilinç için *gerekli* bir koşul mu olduğu (bu Scott için uygundu) yoksa *yeterli* bir koşul mu olduğu (bunu da Giulio ileri sürüyordu) konusunda anlaşamıyor olduklarıydı. İkincisi kesinlikle daha güçlü ve çekişmeli bir iddia, umarım kısa sürede deneysel olarak da test edebiliriz.[23]

Diğer ihtilaflı IIT iddiası da mantık kapılarının bağlanış şekli çok düşük entegrasyon verdiği için bugünün bilgisayar mimarilerinin bilinçli olamayacağıdır.[24] Diğer bir deyişle, kendinizi her bir nöron ve sinapsınızı doğru biçimde simüle eden, geleceğin yüksek güçlü bir robotuna yüklerseniz, bu dijital klon sizden ayırt edilemez şekilde konuşsa, görünse ve davransa bile, Giulio'ya göre öznel deneyimi olmayan bilinçsiz bir zombi olacaktır. Bu da kendinizi öznel ölümsüzlük için yüklediyseniz oldukça büyük bir hayal kırıklığı olacaktır.* Bu iddiaya hem David Chalmers hem de YZ profesörü Murray Shanahan, beyninizdeki sinirsel devreleri, onları mükemmel biçimde simüle eden farazi bir dijital donanımla yavaş yavaş değiştirirsek ne olacağını düşünerek meydan okumuştur.[25] Simülasyon, varsayım gereği mükemmel

* Bu iddia ve bilincin maddeden bağımsız olduğu fikri arasında potansiyel bir gerilim vardır çünkü bilgi işleme en düşük seviyede farklı olsa bile tanımı gereği davranışı belirlediği yüksek seviyelerde özdeştir.

olduğu için değişimden *davranışınız* etkilenmese dahi, Giulio'ya göre *bilinciniz* bilinçliden bilinçsize dönüşecektir. Ancak giderek parçalarınız değiştirilirken arada nasıl hissedersiniz? Görüş alanınızın üst kısmının bilinçli deneyiminden sorumlu beyin kısımları değiştirilirken, görüş alanınızın o kısmının birden kaybolduğunu fark eder ve "kör görüş" hastalarının söylediği gibi, gizemli bir biçimde orada ne olduğunu yine de bilebilir misiniz?[26] Bu oldukça sıkıntılı olur çünkü eğer bilinçli biçimde bir fark deneyimlerseniz, arkadaşlarınıza da sorulduğunda söyleyebilirsiniz ama varsayım gereği, davranışınız değişemez. Bu varsayımlarla uyumlu tek mantıksal olasılık, tam olarak bilincinizden herhangi bir şeyin kaybolduğu anda zihninizin gizemli biçimde ya yalan söylemenizi ya da deneyiminizin değiştiğini reddetmenizi sağlayacak veya şeylerin önceden farklı olduğunu size unutturacak şekilde değişmesidir.

Öte yandan, Murray Shanahan aynı yavaşça değişim eleştirisinin bilinçli olmadan bilinçli davranabileceğinizi iddia eden *herhangi bir* teoriye de yöneltilebileceğini; bu yüzden, bilinçli davranmak ve olmanın aynı şeyler olduğunu ve haricî olarak gözlemlenebilir davranışın önemli olan her şey olduğunu söylemeye teşvik edilebileceğinizi kabul ediyor. Ama sonrasında da öyle olmadığını bilseniz bile, rüya görürken bilinçsiz olduğunuzu tahmin etme tuzağına düşmüş olabilirsiniz.

Üçüncü bir ihtilaflı IIT konusu da bilinçli varlığın kendi içlerinde bilinçli olan parçalardan oluşup oluşmadığıdır. Mesela, insanlar kendi bilinçlerini kaybetmeden toplum bir bütün olarak bilinç kazanabilir mi? Bilinçli bir beynin kendi başlarına bilinç sahibi olan kısımları olabilir mi? IIT'nin tahmini net bir "hayır"dır ama herkes ikna olmuş gibi gözükmüyor. Mesela, beynin iki yarısı arasında iletişimi ciddi biçimde azaltan lezyonlara sahip bazı hastalar "yabancı el sendromu" deneyimlerler. Bu sendromda, sağ

beyinleri sol eline hastaların kendi yaptırmadığı ya da anlamadığı şeyler yaptırır. Bazen bu öyle bir seviyeye çıkar ki diğer elleriyle "yabancı" ellerini durdurmaları gerekir. Kafalarında iki adet farklı bilinç olmadığından nasıl emin olabiliriz? Sağ yarıkürelerinde konuşamayan bir tane, sol yarıkürelerinde ise konuşabilen ve ikisi için de konuştuğunu iddia eden başka biri olamaz mı? Gelecek teknolojisini iki insan beyni arasında doğrudan iletişim kurmak için kullandığımızı ve bu bağlantının kapasitesini, iletişimin beyinlerin kendi içlerinde olduğu kadar beyinler arasında da verimli olacak duruma gelene kadar artırdığımızı hayal edin. İki tekil bilincin birden kaybolduğu ve IIT'nin tahmin ettiği üzere tek birleşmiş bir bilinçle yer değiştirdiği bir an gelecek mi? Yoksa dönüşüm kademeli olacak ve bir biçimde, hatta birleşik deneyim olarak bir arada var olan tekil bilinçler mi ortaya çıkmaya başlayacak?

Bir diğer baş döndürücü ihtilaf ise deneylerin ne kadar bilinçli olduğumuzu hafife alıp almadığı üzerinedir. Daha önceden renkler, şekiller, nesneler ve görünüşte önümüzde duran her şeye dair engin miktarda bilginin görsel olarak bilincinde olduğumuzu *hissetsek* de deneyler, bunların yalnızca çok küçük bir kısmını hatırlayıp nakledebileceğimizi göstermiştir.[27] Bazı araştırmacılar bu çelişkiyi, "erişim olmadan bilince", yani gelecek kullanımlar için çalışan belleğimize sığmayacak kadar karmaşık olan şeylerin öznel deneyimine sahip olup olmadığımızı sorarak çözmeye çalışmışlardır.[28] Mesela, açık bir biçimde önünüzde duran bir nesneyi göremeyecek kadar dikkatiniz dağınıkken *istem dışı körlük* yaşadığınızda, bu hiçbir bilinçli görsel deneyime sahip olmadığınızı değil, yalnızca çalışan hafızanızda depolanmamış olduğunu gösterir.[29] Bu körlükten ziyade unutkanlık olarak mı görülmeli? Başka araştırmacılar, insanların deneyimlediklerini söyledikleri şey hakkında güvenilemeyeceği fikrini reddetmişler ve sonuçları konusunda uyarmışlardır. Murray Shanahan, hastaların yeni

bir mucize ilaç sayesinde tamamen ağrılarından kurtulduklarını söyledikleri bir klinik deneyi tahayyül ediyor. Fakat ilaç yine de hükümet tarafından reddediliyor: "Hastalar yalnızca ağrıları olmadığını düşünüyorlar. Sinirbilim sayesinde böyle olmadığını biliyoruz."[30] Öte yandan, yanlışlıkla ameliyat esnasında uyanan hastalara bu korkunç durumu unutmaları için ilaç verildiği vakalar da olmuştur. Sonrasında hiç acı çekmediklerini söylediklerinde onlara güvenmeli miyiz?[31]

YZ Bilinci Nasıl Hissettirebilir?

Eğer gelecekte bir YZ sistemi bilinçli olursa, öznel olarak ne deneyimleyecek? Bu, bilincin "daha da zor probleminin" özüdür ve Görsel 8.1'de gösterilmiş zorluk seviyesinin ikinci katına çıkmaya zorlar bizi. Bu soruya cevap verecek bir teoriye sahip olmadığımız gibi, buna tamamen cevap vermenin mantıksal olarak mümkün olup olmadığını bile bilmiyoruz. Sonuçta tatminkâr bir cevap nasıl bir şey olurdu ki? Kör doğan bir insana kırmızı rengin ne olduğunu nasıl açıklayabilirsiniz?

Neyse ki şu anda bütünlüklü bir cevap verme acizliğimiz bizi kısmi cevaplar vermekten alıkoyamaz. İnsan duyu sistemini çalışan zeki uzaylılar muhtemelen renklerin iki boyutlu bir yüzeyde (görüş alanımız) her noktayla ilişkilenmiş gibi hissettiren *qualia* olduğunu ama seslerin mekânsal olarak sınırlanmış gibi hissettirmediğini ve acının bedenimizin farklı kısımlarıyla ilişkilenmiş gibi hissettirdiğini çıkarsayacaktır. Retinalarımızın üç tür ışığa duyarlı konik hücreye sahip olduğunu keşfedince de, üç ana rengi deneyimlediğimizi ve tüm diğer renk *qualia*'sının da onların birleşiminden oluştuğunu anlayacaktır. Nöronların bilgiyi beyinde iletmesinin ne kadar sürdüğünü ölçerek, saniye başına ondan fazla bilinçli düşünce ya da algıyı deneyimleyeme-

diğimiz ve TV'de saniyede yirmi dört kareyle filmleri izlerken, bunu ardışık durağan görüntüler olarak değil, devamlı hareket olarak algıladığımız sonucuna varacaktır. Adrenalinin dolaşım sistemimize ne kadar hızlı salındığını ve çözülmeden önce ne kadar süre kaldığını ölçerek, saniyeler içerisinde başlayan ve dakikalar süren öfke patlamaları yaşadığımızı tahmin edebileceklerdir.

Benzeri fizik tabanlı argümanları uygulayarak, yapay bilincin nasıl hissedeceğine dair bilgiye dayalı bazı tahminlerde bulunabiliriz. İlk olarak, olası YZ deneyimlerinin uzayı insanların deneyimleyebileceğine kıyasla *devasadır*. Her duyumuz için bir *qualia* sınıfımız var ancak YZ'lerin çok daha fazla türde duyusu ve dâhilî bilgi temsili olabilir, bu yüzden de YZ olmanın insan olmaya benzeyeceği varsayımı tuzağından kaçınmalıyız.

İkinci olarak, beyin boyutunda bir yapay bilinç bize göre saniye başına milyonlarca kat fazla deneyime sahip olabilir çünkü elektromanyetik sinyaller ışık hızında ilerler, yani nöron sinyallerinden milyonlarca kat fazla. Ancak YZ büyüdükçe, Bölüm 4'te gördüğümüz üzere, kısımları arasında bilgi akışının gerçekleşeceği zamanın sağlanması için küresel düşünceleri daha yavaş olur. Bu yüzden de Dünya boyutunda bir "Gaia" YZ'nin saniyede, tıpkı insan gibi, on bilinçli deneyimi olabileceğini ve galaksi boyutunda bir YZ'nin yalnızca yüz bin yılda bir küresel bir düşünceye sahip olabileceğini düşünebiliriz – yani evrenimizin şu ana kadarki tüm tarihi boyunca ancak yüz deneyim! Tıpkı bilinçli zihnimizin göz kırpma refleksini küçük, hızlı ve bilinçsiz bir sisteme havale etmesi gibi, büyük YZ'ler de hızlanmak için hesaplamaları onları halledebilecek en küçük alt sistemlere havale etme konusunda görece karşı koyulamaz bir dürtüye sahip olacaktır. Yukarıda beyinlerimizde bilinçli bilgi işlemenin diğer türlü bilinçsiz olan bir buzdağının zirvesi olduğunu görmüş olsak da büyük, gelecek YZ'leri için durumun daha uçlarda olabileceğini bekleyebiliriz:

Tek bir bilinçleri varsa, içinde gerçekleşen neredeyse tüm bilgi işlemden habersiz olması olasıdır. Dahası, sahip olduğu bilinçli deneyimler oldukça karmaşık olsa bile, küçük kısımlarının hızlı aktivitelerine göre salyangoz hızında olacaklardır.

Bu akla bilinçli bir varlığın parçalarının da bilinçli olup olamayacağına dair yukarıda bahsettiğimiz tartışmayı getirir. IIT olamayacağını söyler, bu da gelecekte astronomik boyutlardaki bir YZ'nin bilinçli olması durumunda neredeyse tüm bilgi işlemesinin bilinçsiz olacağı anlamına gelir. Bu da küçük YZ'lerden oluşan bir uygarlığın iletişim yetilerini tek bir bilinçli kolektif zihnin oluşacağı noktaya kadar geliştirmesi durumunda çok daha hızlı bireysel bilinçlerinin birden yavaşlayacağı anlamına gelir. Eğer IIT tahmini yanlışsa, kolektif zihin bir dizi küçük bilinçli zihinle aynı anda var olabilir. Gerçekten de mikroskobikten kozmiğe her seviyede bilincin iç içe hiyerarşisi hayal edilebilir.

Yukarıda gördüğümüz üzere, insan beynimizdeki bilinçsiz bilgi işleme psikologların "Sistem 1" dedikleri gayretsiz, hızlı ve otomatik düşünme şekliyle bağlantılı gözükür.[32] Mesela, sizin Sistem 1'iniz bilincinize görsel girdi verisinin yüksek oranda kompleks analizinin en yakın arkadaşınızın geldiğini tespit ettiğini, size hesaplamanın nasıl gerçekleştiğine dair hiçbir fikir vermeden, bildirebilir. Eğer sistemler ve bilinç arasındaki bağlantı geçerli olursa, bu terminolojiyi YZ'lere genelleştirmek, tüm hızlı rutin görevleri bilinçsiz alt birimlere havale etmeyi YZ'nin Sistem 1'i olarak adlandırmak çekici olacaktır. YZ'nin zorlu, yavaş ve kontrollü küresel düşünüşü, eğer bilinçliyse, YZ'nin Sistem 2'si olabilir. Biz insanlar da benim "Sistem 0" adını vereceğim şeyi içeren bilinçli deneyimlere sahibiz: hareket etmeden ya da düşünmeden otururken ve sadece etrafınızdaki dünyayı gözlemlerken bile gerçekleşen ham pasif algılama. Sistemler 0, 1 ve 2 giderek

artan biçimde daha karmaşık görünür bu yüzden de yalnızca ortadakinin bilinçsiz olması dikkat çekicidir. IIT bunu, Sistem 0'daki ham duyu bilgisinin oldukça yüksek entegrasyonla ızgara benzeri beyin yapısında depolandığını ancak Sistem 2'nin şu an farkında olduğunuz tüm bilginin gelecek beyin durumunuzu etkileyebileceği geri besleme döngüsünden dolayı yüksek entegrasyona sahip olduğunu söyleyerek açıklar. Öte yandan, Scott Aaronson'ın önceden bahsettiğimiz IIT eleştirisini tetikleyen de tam olarak bu bilinçli ızgara tahminiydi. Özet olarak, eğer bilincin oldukça zor problemini çözebilen bir teori bir gün bir dizi katı deneysel testi geçer ve tahminlerini ciddiye almaya başlarsak, o zaman geleceğin bilinçli YZ'lerinin deneyimleyebileceği daha da zor problem için seçenekleri ciddi oranda azaltabileceğiz.

Öznel deneyimimizin kendimizi korumak (yeme, içme, ölmekten kaçınma) ve üremekle ilişkili duygusal arzularımız gibi bazı veçhelerini açık biçimde evrimsel kökenlerimize kadar takip edebiliriz. Bu da açlık, susuzluk, korku ya da cinsel arzu gibi *qualia*'ları asla deneyimlemeyecek YZ'yi oluşturmanın mümkün olması gerektiği anlamına gelir. Önceki bölümde gördüğümüz üzere, eğer yüksek oranda zeki bir YZ herhangi bir yeterince hırslı hedefe sahip olabilecek şekilde programlanırsa, o hedefe ulaşabilmek için kendini korumaya gayret göstermesi olasıdır. Ancak eğer bir YZ'ler toplumunun parçasıysalar, bizim güçlü insani ölüm korkumuza sahip olmayabilirler: Kendilerini yedekledikleri sürece, kaybedecekleri tek şey –yedeklenmiş yazılımlarının kullanılacağından eminseler- en son yedeklemeden bu yana biriktirdikleri anılar olur. Ek olarak, YZ'ler arasında bilgi ve yazılım kopyalama yetisi insan bilincinin karakteristiklerinden olan güçlü bireysellik duyumunu azaltacaktır: Eğer tüm anılarımızı ve yeteneklerimizi kolayca kopyalayıp paylaşabilirsek, sen

ve ben arasında çok daha az fark kalacaktır, bu yüzden de bir grup yakın YZ daha çok, kolektif zihne sahip tek bir organizma gibi hissedecektir.

Yapay bir bilinç, özgür iradesi olduğunu hissedecek mi? Felsefecilerin, soruyu nasıl tanımaları gerektiğine dair bir fikir birliğine bile varmadan *bizim* özgür iradeye sahip olup olmadığımız konusunda binlerce yıldır tartışıp durmalarına rağmen,[33] benim tartışmalı bir biçimde cevaplanması daha kolay, farklı bir soru sorduğuma dikkatinizi çekmek isterim. Cevabın basit bir biçimde, "Evet, herhangi bir bilinçli karar verici, biyolojik ya da yapay olmasından bağımsız olarak, özgür iradesi olduğunu öznel olarak *hissedecektir*," olduğuna sizi ikna etmeme izin verin. Kararlar iki uç nokta arasında bir yelpazeye düşmektedir:

1. Bu belirli seçimi neden yaptığınızı kesin olarak biliyorsunuz.
2. Bu belirli seçimi neden yaptığınıza dair hiçbir fikriniz yok – bir anlık istekle rastgele seçim yapmışsınız gibi geliyor.

Özgür irade tartışmaları genelde hedefe yönelik karar verme davranışlarımızı fizik yasalarıyla uzlaştırma çabası etrafında döner: Eğer yaptığınız şey için aşağıdaki iki açıklama arasından birini seçiyorsanız, hangisi doğrudur: "*Ona çıkma teklifi ettim çünkü ondan gerçekten hoşlanıyorum*" ya da "*Parçacıklarım fizik yasalarına göre hareket ederek bana bunu yaptırttı*"? Ama son bölümde *ikisinin de* doğru olduğunu gördük: Hedefe yönelik davranış olarak hissedilen, hedefsiz deterministik fizik yasalarına dayanmaktadır. Daha özel olarak, bir sistem (beyin ya da YZ) Tip 1 tarzı bir karar verdiğinde, neye karar vereceğini bazı deterministik algo-

ritmalar kullanarak hesaplar ve buna karar vermiş gibi hissetmesinin nedeni de neyi yapacağını hesaplarken gerçekten de karar vermiş olmasıdır. Dahası, Seth Lloyd tarafından da altı çizildiği üzere,[34] neredeyse tüm hesaplamalar için, onları çalıştırmaktan daha hızlı bir şekilde sonuçlarını tespit etmenin bir yolu olmadığını söyleyen ünlü bir bilgisayar bilimi teoremi vardır. Bu da özgür iradeye sahip olma deneyiminizi güçlendirmeye yardımcı olarak, bir saniye içinde neye karar vereceğinizi bir saniyeden daha kısa sürede bulmanızın aslında imkânsız olduğu anlamına gelir. Aksine, bir sistem (beyin ya da YZ) Tip 2 tarzı bir karar verdiğinde, basit bir biçimde zihnini, rastgele rakam üreticisi gibi hareket eden bir alt sistemin sonucu üzerine kararını verecek şekilde programlar. Beyinlerde ve bilgisayarlarda, etkili rastgele rakamlar yükseltilen gürültüyle kolayca üretilir. Bir kararın 1'den 2'ye spektrumun neresine düşeceğinden bağımsız olarak, hem biyolojik hem de yapay bilinç, özgür iradesi olduğunu hisseder: Karar verenin gerçekten kendileri olduğunu hissederler ve düşünmeyi bitirene dek kararın ne olacağını net biçimde kestiremezler.

Bazı insanlar bana nedenselliği alçaltıcı bulduklarını, düşünme süreçlerini anlamsız kıldığını ve onları "yalnızca" makineye indirgediğini söylüyorlar. Ben böylesi olumsuz bir bakış açısını absürt ve haksız buluyorum. Öncelikle, insan beyniyle ilgili "yalnızca" denebilecek hiçbir şey yok. Bana sorarsanız, insan beyni bilinen evrenimizdeki en inanılmaz ve en sofistike fiziksel nesnedir. İkinci olarak, hangi alternatifi tercih ederlerdi? Kararları verenin kendi düşünce süreçleri (beyinlerinin gerçekleştirdiği hesaplamalar) olmasını istemiyorlar mı? Onların özgür iradelerinin öznel deneyimi, hesaplamalarının içeriden nasıl hissettiğidir: Bitirene kadar bir hesaplamanın sonucunu bilmezler. Bu da hesaplamanın karar *olduğunu* söylemek anlamına gelir.

Anlam

Kitaba başlangıç noktasına dönerek son verelim: Yaşamın geleceğinin nasıl olmasını istiyoruz? Bir önceki bölümde, dünyanın etrafındaki farklı kültürlerin hepsinin pozitif deneyimlerle dolup taşan bir geleceğin peşinde olduğunu ama neyin pozitif olması gerektiği ve farklı yaşam formları için neyin iyi olduğu üzerinden nasıl değiş tokuş yapılması gerektiği üzerine fikir birliği ararken baş döndürücü derecede çetrefilli ihtilafların ortaya çıktığını gördük. Fakat şimdi bu ihtilafların bizi gözümüzün önündeki gerçeği görmekten alıkoymasına izin vermeyelim: Eğer hiç deneyim olmazsa, yani hiç bilinç olmazsa, hiçbir pozitif deneyim de olamaz. Diğer bir deyişle, bilinç olmadan, mutluluk, iyilik, güzellik anlam ya da amaç olamaz; elimizde olan tek şey astronomik bir yer israfıdır. Bu da insanlar, sanki var oluşumuza anlam vermek kozmosumuzun işiymişçesine yaşamın anlamını sorduğunda, her şeyi tersten anladıklarına işaret eder: *Bilinçli varlıklara anlam veren evrenimiz değildir, asıl bilinçli varlıklar evrenimize anlam verir.* O zaman gelecek için hazırladığımız dilek listesinde ilk hedefimiz biyolojik ve/veya yapay bilinci yok etmekten ziyade, kozmosumuzda kalmasını (ve umalım ki genişlemesini) sağlamak olmalıdır.

Eğer bu çabamızda başarılı olursak, insanlar kendilerinden daha zeki makinelerle bir arada yaşamak konusunda ne düşünecekler? Yapay zekânın görünüşte aman vermez yükselişi sizi rahatsız ediyor mu ve ediyorsa, neden? Bölüm 3'te YZ destekli teknolojinin güvenlik ve gelir gibi temel ihtiyaçlarımızı karşılamasının görece kolay olduğunu, tek gerekenin bunları karşılamak için politik iradenin de bulunması olduğunu görmüştük. Ancak iyi beslenme, giyinme, başınızı sokacak bir evinizin olması ve biraz eğlenmenin yeterli olmayacağını düşünüyorsunuz muhtemelen. Eğer YZ'nin tüm pratik ihtiyaçlarımız ve arzularımızı karşılayacağı konusunda bize garanti verilirse, iyi bakılan hay-

vanat bahçesi hayvanları gibi, yine de yaşamlarımızda anlam ve amacın eksikliğini hisseder miyiz?

Geleneksel olarak, biz insanlar kendimize verdiğimiz öz değeri *insan özgücülük* fikri üzerine kurarız: Gezegendeki en zeki varlıklar olduğumuz ve bu yüzden de eşsiz ve üstün olduğumuz inancı. YZ'nin yükselişi bizi bundan vazgeçmeye ve daha alçak gönüllü olmaya zorlayacak. Ancak belki de bu zaten yapmamız gereken bir şey: Sonuçta, diğerlerinden (bireylerden, etnik gruplardan, türlerden vb.) üstün olduğumuz kibrine tutunmak geçmişte korkunç sorunlara yol açmıştır ve belki de artık emekliliğe ayrılması gereken bir fikirdir. Gerçekten de insan özgücülük geçmişte yalnızca kedere yol açmakla kalmamış, insan neslinin gelişmesi için de gereksizdir: Eğer barışçıl bir dünya dışı uygarlığın bilim, sanat ve umursadığımız her şeyde bizden daha ileride olduğunu keşfedersek, bunun yaşamlarında anlam ve amaç deneyimlemeyi sürdürmekten insanları alıkoymaması gerekir. Ailelerimizi, dostlarımızı ve daha büyük topluluklarımız ile bize anlam ve amaç veren tüm aktivitelerimizi korurken, kibir dışında başka hiçbir şeyi kaybetmeyiz.

Geleceğimizi planlarken, yalnızca kendi yaşamlarımız için değil evrenimizin kendisi için de anlamı değerlendirelim. Bu noktada en beğendiğim fizikçilerimden ikisi, Steven Weinberg ve Freeman Dyson taban tabana zıt görüşler sergiliyorlar. Parçacık fiziğinin standart modeli üzerine yaptığı esaslı çalışmayla Nobel Ödülü kazanan Weinberg, şu meşhur lafı söylemiştir: "Evren ne kadar anlaşılabilir gözükürse, o kadar da anlamsız gözükür."[35] Öte yandan Dyson, Bölüm 6'da gördüğümüz üzere çok daha optimisttir: Evrenimizin anlamsız *olduğuna* katılsa da yaşamın onu anlamla doldurduğuna, en iyisinin de yaşamın kozmosa yayıldığı zaman geleceğine inanır. 1979 tarihli ufuk açıcı makalesini de bu yüzden şöyle bitirmiştir: "Weinberg'in evreni mi yoksa benimki

mi hakikate daha yakın? Bir gün, hem de çok uzak bir zamanda değil, bunu öğreneceğiz."[36] Eğer evrenimiz dünya yaşamının soyunu tükettiğimiz ya da bilinçsiz zombi YZ'nin evrenimizi ele geçirmesine izin verdiğimiz için temelli bilinçsiz olmaya geri dönerse, o zaman Weinberg fazlasıyla haklı çıkmış olacak.

Bu perspektiften, bu kitapta geleceğin zekâsına odaklanmış olsak da geleceğin bilincinin çok daha önemli olduğunu görüyoruz çünkü anlamı mümkün kılan odur. Filozoflar bu ayrımı Latince çözmeye çalışırlar, *sapience* (zekice düşünme kabiliyeti) ile *sentience*'ı (öznel olarak *qualia*'yı deneyimleme yetisi) karşı karşıya getirirler. Biz insanlar kimliğimizi *Homo Sapiens* olmak üzerine kurduk, yani etrafta bulunan en zeki varlıklar. Bizden daha zeki makineler tarafından ikinci plana atılmaya hazırlanırken, kendimizi *Homo Sentiens* olarak yeniden adlandırmayı öneriyorum!

SONUÇ:

- "Bilincin" tartışmasız bir tanımı yoktur. Ben geniş ve insan merkezci olmayan tanımı kullanıyorum. *Bilinç = Öznel deneyim.*
- YZ'lerin o şekilde bilinçli olup olmadığı YZ'nin yükselişiyle ortaya çıkan en çetrefilli etik ve felsefi problemler için önemli olan şeydir: YZ'ler acı çekebilir mi? Hakları olmalı mı? Yükleme öznel bir intihar mıdır? YZ'lerle dolup taşan geleceğin kozmosu asıl zombi kıyameti mi olacak?
- Zekâyı anlama problemi üç ayrı bilinç problemiyle birleştirilmemelidir: Hangi fiziksel sistemin bilinçli olduğunu tahmin etmeyi kapsayan "oldukça zor problem", *qualia*'yı tahmin etmeyi kapsayan "daha da zor problem" ve herhangi bir şeyin neden bilinçli olduğunu soruşturan "*gerçekten* zor problem".

- Bilincin "oldukça zor problemi" bilimseldir çünkü beyin süreçlerinin hangilerinin bilinçli olup olmadığını tahmin eden bir teori deneysel olarak test edilebilir ve yanlışlanabilirken, bilimin en zor iki soruyu tamamen nasıl çözebileceği halen belirsizdir.
- Sinirbilim deneyleri pek çok davranış ve beyin bölgesinin bilinçsiz olduklarını, bilinçli deneyimimizin büyük kısmının geniş miktarda bilinçsiz bilginin olay sonrası özetini temsil ettiğini ortaya koyar.
- Bilinç tahminlerini beyinlerden makinelere genellemek bir teori gerektirir. Bilinç belirli bir tür parçacık ya da alan değil, oldukça özerk ve entegre belirli bir tür bilgi işlemeye ihtiyaç duyar gibi görünür, böylece tüm sistem görece özerktir ama parçaları değildir.
- Bilinç fiziksel değilmiş gibi hissedilir çünkü iki kere maddeden bağımsızdır: Eğer bilinç, çeşitli karmaşık şekillerde işlenirken bilginin hissettiği ise, önemli olan şey bilgi işlemenin yapısıdır, bilgi işlemeyi yapan maddenin yapısı değil.
- Eğer yapay bilinç mümkünse, mümkün YZ deneyimleri uzayı, biz insanların deneyimleyebileceğine kıyasla devasadır ve geniş bir *qualia* ile zaman eksenleri spektrumuna yayılır. Hepsi de özgür iradeye sahip olma duygusunu paylaşır.
- Bilinç olmadan anlam olamayacağı için, bilinçli varlıklara anlam veren evrenimiz değildir, bilinçli varlıklar evrenimize anlam verir.
- Bu da giderek zekileşen makinelerce ikinci plana atılmayı beklerken, *Homo sapiens* değil de *Homo sentiens* olmakta avuntu bulabileceğimiz anlamına gelir.

Sonsöz:



FLI Takımının Hikâyesi

Şu an yaşamın en hüzün verici yanı, bilimin bilgi toplama hızının toplumun bilgelik toplama hızından daha fazla olmasıdır.

Isaac Asimov

İşte sevgili okuyucum, zekânın, hedeflerin ve anlamın kökeni ve kaderini inceledikten sonra kitabın sonuna geldik. Bu fikirleri nasıl harekete dönüştürebiliriz? Geleceğimizi mümkün olduğunca iyileştirmek için somut olarak ne *yapmalıyız*? Asilomar'da düzenlediğimiz YZ Konferansı sonrasında, 9 Ocak 2017'de San Francisco'dan Boston'a dönerken pencere kenarındaki koltuğumda şu an otururken kendime sorduğum soru da tam olarak bu, o yüzden fikirlerimi sizinle paylaşarak kitabı bitirmeme izin verin.

Meia hazırlık ve organizasyonla dolu kısa gecelerde ihmal ettiği uykusunu şu an yanımda uyuyarak telafi ediyor. Vay be... Ne çılgın bir haftaydı! Porto Riko'nun devamında, bu kitapta bahsettiğim neredeyse tüm insanları birkaç günlüğüne bir araya getirmeyi başardık. Aralarında Elon Musk ve Larry Page gibi girişimciler ile DeepMind, Google, Facebook, Apple, IBM, Microsoft ve Baidu gibi firmalar ve akademiden YZ araştırma liderleri, onlara ek olarak da ekonomistler, hukukçular, felsefeciler ve diğer muazzam düşünürler var (bkz. Görsel 9.1). Sonuçlar benim

yüksek beklentilerimi bile aştı ve yaşamın geleceği hakkında uzun süredir olmadığım kadar iyimser hissediyorum. Bu sonsözde size neden olduğunu söyleyeceğim.

FLI Doğdu

On dört yaşımda nükleer silah yarışını duyduğumdan beri, teknolojimizin gücünün onu yönettiğimiz bilgelikten daha hızlı büyüyor olmasından endişe duyarım. Bu sebeple, tamamen fizik üzerine olsa da ilk kitabım *Our Mathematical Universe*'e bu zorluk üzerine bir bölüm koymaya karar verdim. 2014 yılında bir yeni yıl kararı verdim. Artık kişisel olarak ne yapabileceğim üzerine ciddi ciddi düşünmeden hiçbir konuda şikâyet etmeyecektim, ocak ayında kitap turum sırasında da bu sözümü tuttum: Meia ve ben teknolojik idare ile yaşamın geleceğini iyileştirmeye odaklanmış bir tür kâr amacı gütmeyen kurum üzerine beyin fırtınası yaptık.

Eşim, bu kuruma "Kıyamet & Kasvet Enstitüsü" ve "Hadi Gelecek Hakkında Endişe Edelim Enstitüsü" gibi isimlerden olabildiğince farklı, olumlu bir isim vermemiz gerektiği konusunda ısrarcıydı. İnsanlığın Geleceği Enstitüsü ismi çoktan alındığı için, daha kapsayıcı olma avantajına da sahip Yaşamın Geleceği Enstitüsü / Future of Life Institute (FLI) isminde karar kıldık. 22 Ocak'ta kitap turunda Santa Cruz'a geldim ve California güneşi Pasifik üzerinde batarken, eski dostumuz Anthony Aguirre'la bir akşam yemeği yedik ve bize katılması için onu ikna ettik. Anthony yalnızca tanıdığım en bilge ve idealist insanlardan biri değil, ayrıca benimle on yılı aşkın süredir başka bir kâr amacı gütmeyen kurumu, *Foundational Questions Institute / Temel Sorular Enstitüsü*'nü (bkz. http://fqxi.org) yürütmeye katlanmayı başaran birisidir de.

Ertesi hafta, tur beni Londra'ya götürdü. YZ'nin geleceği hep aklımda olduğu için, beni DeepMind'ın merkezini ziyaret etmem için nazikçe davet eden Demis Hassabis'e ulaştım. İki yıl önce beni MIT'de ziyaret ettiğinden beri ne kadar çok büyüdüklerini görmek beni çok şaşırttı. Google onları yaklaşık altı yüz elli milyon dolara almıştı ve Demis'in gözüpek "zekâyı çözmek" hedefini takip eden parlak zekâlarla dolu devasa ofisi görmek bana başarının gerçekten mümkün olduğuna dair içgüdüsel bir his verdi.

Ertesi akşam, arkadaşım Jaan Tallinn'le oluşturulmasına yardım ettiği yazılım Skype'tan konuştum. FLI vizyonumuzu açıkladım ve bir saat sonra, bize bir şans vererek, bizi yılda yüz bin dolara kadar çıkabilecek şekilde fonlamaya karar verdi! Dünyada pek az şey beni hak ettiğimden daha çok bana güvenilmesi kadar etkiler, bu yüzden de bir yıl sonra, Bölüm 1'de bahsettiğim Porto Riko Konferansı'nın ertesinde, bunun yaptığı en iyi yatırım olduğu konusunda şaka yapması benim için çok anlamlıydı.

Sonraki gün, yayıncım takvimimde bir boşluk bıraktı, ben de bunu Londra Bilim Müzesi'ne giderek değerlendirdim. Uzun yıllar boyunca zekânın geçmişi ve geleceğini takıntı haline getirdikten sonra, birden düşüncelerimin fiziksel tezahürüne doğru yürüdüğümü fark ettim. Stephenson'ın roket lokomotifinden Ford Model T'ye, gerçek boyutlu Apollo 11 aya iniş aracı kopyasına ve Babbage'ın "Fark Makinesi" isimli mekanik hesap makinesinden günümüz donanımlarına kadar pek çok bilgisayara dek, bilgimizin büyümesini temsil eden şeylerin fantastik bir koleksiyonunu bir araya getirmişlerdi. Ayrıca, içinde Galvano'nun kurbağa bacağı deneyinden nöronlar, EEG ve fMRI'a kadar zihni kavrayışımızın tarihi hakkında bir sergileri de vardı.

Çok nadiren ağlarım ama çıkarken ağladım; hem de yayalarla dolu bir tünelde South Kensington metro istasyonuna giderken.

Etrafımda ne düşündüğümden habersiz, hayatlarını mutlu mutlu yaşayan bir sürü insan vardı. İlk olarak biz insanlar makinelerle bazı doğal süreçleri nasıl kopyalayacağımızı keşfederek kendi rüzgâr ve yıldırımımızı, kendi mekanik beygir gücümüzü yaptık. Yavaş yavaş bedenlerimizin de makineler olduğunu anlamaya başladık. Ardından sinir hücrelerinin keşfi beden ve zihin arasındaki sınırı bulanıklaştırmaya başladı. Sonrasında da sadece kaslarımıza değil zihnimize de üstün gelen makineler inşa etmeye başladık. Kim olduğumuzu keşfetmemize paralel olarak, kendimizi kaçınılmaz bir biçimde hükümsüz mü kılıyorduk? Bu şiirsel bir biçimde trajik olurdu.

Bu düşünce beni korkuttu ancak yeni yıl kararıma sıkı sıkıya bağlı kalmamı sağladı. İdealist ve genç gönüllülerden oluşan bir takımın başını çekecek bir kişiye daha ihtiyacımız olduğunu, böylece FLI kurucular takımını tamamlayabileceğimizi hissettim. Mantıklı seçim Viktoriya Krakovna olacaktı. Krakovna, Uluslararası Matematik Olimpiyatları'nda bir gümüş madalya kazanmış, yaşamlarında ve dünyada daha büyük bir rol oynamak için sebepler arayan bir düzine genç idealistin içinde bulunduğu Citadel'i kurmuş, parlak bir Harvard doktora öğrencisiydi. Meia ve ben beş gün sonra onu eve davet ettik, ona vizyonumuzu anlattık ve suşimizi bitirmeden evvel FLI doğmuştu.



Görsel 9.1: 2017 Ocak ayında Porto Riko Konferansı'nın devamı olarak düzenlediğimiz Asilomar Konferansı, YZ ve alakalı alanlardan bir grup olağanüstü araştırmacıyı bir araya getirdi. Arka sıra, soldan sağa: Patrick Lin, Daniel Weld, Ariel Conn, Nancy Chang, Tom Mitchell, Ray Kurzweil, Daniel Dewey, Margaret Boden, Peter Norvig, Nick Hay, Moşe Vardi, Scott Siskind, Nick Bostrom, Francesca Rossi, Shane Legg, Manuela Veloso, David Marble, Katja Grace, Irakli Beridze, Marty Tenenbaum, Gill Pratt, Martin Rees, Joshua Greene, Matt Scherer, Angela Kane, Amara Angelica, Jeff Mohr, Mustafa Suleyman, Steve Omohundro, Kate Crawford, Vitalik Buterin, Yutaka Matsuo, Stefano Ermon, Michael Wellman, Bas Steunebrink, Wendell Wallach, Allan Dafoe, Toby Ord, Thomas Dietterich, Daniel Kahneman, Dario Amodei, Eric Drexler, Tomaso Poggio, Eric Schmidt, Pedro Ortega, David Leake, Seán Ó hÉigeartaigh, Owain Evans, Jaan Tallinn, Anca Dragan, Sean Legassick, Toby Walsh, Peter Asaro, Kay Firth-Butterfield, Philip Sabes, Paul Merolla, Bart Selman, Tucker Davey, ?, Jacob Steinhardt, Moşe Looks, Josh Tenenbaum, Tom Gruber, Andrew Ng, Kareem Ayoub, Craig Calhoun, Percy Liang, Helen Toner, David Chalmers, Richard Sutton, Claudia Passos-Ferriera, János Krámar, William MacAskill, Eliezer Yudkowsky, Brian Ziebart, Huw Price, Carl Shulman, Neil Lawrence, Richard Mallah, Jurgen Schmidhuber, Dileep George, Jonathan Rothberg, Noah Rothberg. Ön sıra: Anthony Aguirre, Sonia Sachs, Lucas Perry, Jeffrey Sachs, Vincent Conitzer, Steve Goose, Viktoriya Krakovna, Owen Cotton-Barratt, Daniela Rus, Dylan Hadfield-Menell, Verity Harding, Shivon Zilis, Laurent Orseau, Ramana Kumar,



Nate Soares, Andrew McAfee, Jack Clark, Anna Salamon, Long Ouyang, Andrew Critch, Paul Christiano, Yoshua Bengio, David Sanford, Catherine Olsson, Jessica Taylor, Martina Kunz, Kristinn Thorisson, Stuart Armstrong, Yann LeCun, Alexander Tamas, Roman Yampolskiy, Marin Soljačić, Lawrence Krauss, Stuart Russell, Eric Brynjolfsson, Ryan Calo, ShaoLan Hsueh, Meia Chita-Tegmark, Kent Walker, Heather Roff, Meredith Whittaker, Max Tegmark, Adrian Weller, Jose Hernandez-Orallo, Andrew Maynard, John Hering, Abram Demski, Nicolas Berggruen, Gregory Bonnet, Sam Harris, Tim Hwang, Andrew Snyder-Beattie, Marta Halina, Sebastian Farquhar, Stephen Cave, Jan Leike, Tasha McCauley, Joseph Gordon-Levitt. Sonradan gelenler: Guru Banavar, Demis Hassabis, Rao Kambhampati, Elon Musk, Larry Page, Anthony Romero.

Porto Riko Macerası

Bu halen süren inanılmaz bir maceranın başlangıcıydı. Bölüm 1'de bahsettiğim gibi, evimizde düzinelerce idealist öğrenci, profesör ve diğer düşünürlerle, üst düzey fikirlerin projelere dönüştüğü düzenli beyin fırtınası toplantıları düzenliyorduk. Bu bahsettiğim projelerden ilki de Bölüm 1'de de adı geçen, Stephen Hawking, Stuart Russell ve Frank Wilczek'le birlikte yazdığımız ve genel

bir tartışmayı alevlendiren YZ hakkında serbest kürsü yazısıydı. Yeni bir kurum kurmanın ilk adımlarıyla (bir danışma kurulu kurmak ve web sitesi açmak gibi) paralel olarak, Alan Alda'nın teknolojinin geleceğini önde gelen uzmanlarla keşfettiği, kalabalık bir MIT salonunda eğlenceli bir lansman etkinliği düzenledik.

Yılın kalan kısmında Bölüm 1'de bahsettiğim üzere YZ'yi faydalı tutmak için neler yapılması gerektiği tartışmasına dünyanın önde gelen YZ araştırmacılarını dâhil etmeyi hedeflediğim Porto Riko Konferansı'nı yapmaya çalıştık. Hedefimiz YZ güvenliği tartışmasını, oturup endişe etmekten ziyade kalkıp çalışmaya çevirmekti: Ne kadar endişeli olmamız gerektiği hakkında çekişmekten, iyi bir sonuç elde etme şansını maksimize etmek için hemen başlatılabilecek somut araştırma projeleri üzerine anlaşmaya. Hazırlanmak için, tüm dünyadan gelecek vadeden YZ güvenliği araştırma fikirlerini topladık ve büyüyen proje listemiz üzerine camiadan geri bildirim almaya çalıştık. Stuart Russell'ın ve bir grup genç gönüllünün –özellikle de Daniel Dewey, Janos Kramar ve Richard Mallah'ın– yardımlarıyla, bu araştırma önceliklerini konferansta tartışılacak bir dokümanda toparladık.[1] Yapılabilecek pek çok değerli YZ güvenliği araştırması olduğu üzerine fikir birliği inşa etmenin insanları böyle araştırmalar yapmaya teşvik edeceğini umuyorduk. Olabileceklerin en iyisi de birinin bunu finanse etmeye ikna olması olacaktı çünkü o ana dek hiçbir hükümet kanalı böyle bir işe destek olmamıştı.

Derken Elon Musk devreye girdi. 2 Ağustos'taki şu tweet'i ile radarımıza girdi: "Bostrom'un *Superintelligence*'ını okumanız gerek. YZ konusunda çok dikkatli olmalıyız. Potansiyel olarak nükleerden bile daha tehlikeli." Ona ulaşıp çabalarımızı anlattım ve birkaç hafta sonra telefonda konuştuk. Çok gergin ve ünlü biriyle tanışmanın heyecanı içinde olsam da sonuç çok iyiydi: FLI bilimsel danışma kurulumuza katılmaya, konferansımıza

gelmeye ve Porto Riko'da açıklanacak ilk YZ güvenliği araştırma programını potansiyel olarak finanse etmeye razı olmuştu. Bu FLI'deki herkesi heyecanlandırdı ve inanılmaz bir konferans yapmak, gelecek vadeden araştırma konuları bulmamız ve onlar için camiadan destek inşa etmek için çabalarımızı ikiye katlamamızı sağladı.

İki ay sonra bir uzay sempozyumu için MIT'ye geldiğinde geleceği planlamak için sonunda Elon'la tanıştım. Bir rock yıldızı gibi binden fazla MIT öğrencisinin aklını başından aldıktan sonra o küçük yeşil odada onunla yalnız kalmak garip gelmişti ama birkaç dakika sonra, düşünebildiğim tek şey ortak projemizdi. Onu hemen sevdim. Samimiydi ve insanlığın uzun vadeli geleceğini çok önemsemesi beni etkilemişti. Hem de gözüpek biçimde hedeflerini gerçeğe dönüştürüyordu. İnsanlığın evrenimizi keşfedip oraya yerleşmesini istiyordu, bu yüzden de bir uzay şirketi kurmuştu. Sürdürülebilir enerji istiyordu, o yüzden de bir güneş enerjisi şirketi ve elektrikli araba şirketi kurmuştu. Uzun, yakışıklı, belagat sahibi ve oldukça bilgiliydi, insanların onu neden dinlediğini anlamak kolaydı.

Maalesef bu MIT etkinliği bana medyanın ne kadar korkuya dayalı ve anlaşmazlık çıkarıcı olduğunu da öğretti. Elon'un sahne performansı uzayın keşfi hakkında baş döndürücü bir saatlik bir tartışmaydı. Bence TV'de de çok izlenirdi. Sonlara doğru bir öğrenci YZ hakkında alakasız bir soru sordu. Elon'un cevabında, "yapay zekâyla şeytanı çağırıyoruz" şeklinde bir kısım vardı. Bu da çoğu medyanın haber yaptığı *tek* şey oldu; hem de bağlam dışı bir biçimde. Pek çok gazetecinin Porto Riko'da yapmaya çalıştığımız şeyin *tam tersini* yanlışlıkla yapıyor olması şaşırtıcıydı. Ortak noktaların altını çizerek genel bir kamuoyu yaratmaya çalışırken medya da farklılıkların altını çizmeye çalışıyordu. Ne kadar fazla ihtilaf yansıtırlarsa, o kadar fazla Nielsen reytingi ve

reklam geliri alacaklardı. Dahası, biz farklı insanların bir araya gelmesi için çaba gösterirken medya farklı fikirlere sahip insanların birbirinin canını sıkmasına yol açıyor, bağlamdan kopuk biçimde en provokatif sözlerini yayımlayarak yanlış anlaşılmaları körüklüyordu. Bu sebeple, Porto Riko konferansına gazetecileri almamaya ve katılımcıların kimin ne dediğini açık etmesini yasaklayan "Chatham House Kuralı'nı" uygulamaya karar verdik.*

Porto Riko Konferansı'mız başarılı olsa da kolay olmadı. Son günlerde gayretli bir hazırlık işi vardı. Mesela ben bir dizi YZ araştırmacısını arayıp ya da Skype'la onlarla konuşup diğer katılımcıları etkilemeleri için yeterli çoğunluk kurmaya çalışıyordum ve dramatik anlar da oluyordu; kötü bir telefon bağlantısı ile Uruguay'da Elon'a ulaşmak için 27 Aralık günü sabah 7'de kalkmam ve bana, "Bunun işe yarayacağını sanmıyorum," denmesi gibi. YZ güvenlik araştırma programının yanlış bir güvenlik hissi yaratacağını, güvenliği umursuyormuş gibi yapan bazı umursamaz araştırmacıların hızla ilerlemesine neden olacağını düşünüyordu. Ancak sesin devamlı kesilmesine rağmen, bu konuyu gündeme taşımanın ve daha fazla YZ araştırmacısının YZ güvenliği üzerine çalışmasını sağlamanın büyük faydaları konusunda derinlemesine konuştuk. Hat kesildikten sonra, bana en sevdiğim e-postalarımdan birini gönderdi: "Hat kesildi burada. Neyse, dosyalar iyi gözüküyor. Bu araştırmayı üç yıllığına beş milyon dolarla desteklemekten mutlu olurum. Belki de on milyon yapmalıyız, ne dersin?"

Dört gün sonra, 2015 Meia ve benim için güzel bir şekilde başladı. Kısa süreliğine, toplantı öncesi dinlendik ve yeni yılda havai fişeklerin aydınlattığı bir Porto Riko sahilinde dans ettik. Konferans da çok güzel başladı: Daha fazla YZ güvenliği

* Bu deneyim kişisel olarak haberleri nasıl yorumlamam gerektiği konusunda tekrar düşünmemi sağladı.

araştırmasına ihtiyaç duyulduğu üzerine ciddi bir birlik vardı ve konferans katılımcılarından gelen yorumlarla birlikte, üzerine çok çalıştığımız araştırma öncelikleri dokümanı iyileştirildi ve son halini aldı. Bölüm 1'deki güvenlik araştırmasını destekleyen açık mektubu dağıttık ve neredeyse herkesin imzalamasından dolayı çok mutlu olduk.

Meia ve ben otel odamızda Elon'la büyülü bir toplantı yaptık. Destek programımızın detaylı planını kabul etti. Meia, kişisel hayatına dair gerçekçiliği ve açıklığıyla bize gösterdiği ilgiden etkilenmişti. Bize nasıl tanıştığımızı sordu ve Meia'nın ayrıntılı hikâyesini beğendi. Ertesi gün, YZ güvenliği ve bunu neden desteklemek istediği hakkında onunla bir röportaj çektik, her şey yolunda gözüküyordu.[2]

Konferansın doruk noktası, Elon'un bağış açıklaması, 4 Ocak 2015 Pazar günü akşam 7'de olacaktı. O kadar gergindim ki önceki gece bir o yana bir bu yana dönmekten uyuyamadım. Sonrasında da bu açıklamanın gerçekleşmesi gereken oturuma gitmeden on beş dakika önce, bir çıkmaza girdik! Elon'un asistanı aradı ve Elon'un açıklamayı yapamayabileceğini söyledi. Meia beni hiç bu kadar stresli ya da hayal kırıklığına uğramış görmediğini söyledi. Neyse ki Elon geldi ve orada oturup konuşurken, oturumun başlamasını beklerken saniyelerin akıp gitmesini izliyordum. Bir gemiye ilk başarılı inişi yapmayı umut ettikleri hayati önem taşıyan bir SpaceX roket fırlatılışına yalnızca iki gün kaldığını ve bu çok önemli bir kilometre taşı olduğu için, SpaceX ekibinin eş zamanlı bir gündemin onun dikkatini dağıtmasını istemediğini söyledi. Her zamanki gibi serinkanlı ve mantıklı bir biçimde Anthony Aguirre, bunun ne Elon'un ne de YZ camiasının medya ilgisini istemediği anlamına geldiğinin altını çizdi. Moderasyonunu yaptığım oturuma birkaç dakika geç vardık ama bir planımız vardı: Hiçbir şekilde miktardan bahsedilmeyecekti, böylece haber niteliği

ortadan kalkacaktı ve ben de Chatham House Kuralı'nı uygulayacak, eğer roket hedefine ulaşırsa, inişin başarıya ulaşmasından bağımsız olarak herkesin Elon'un açıklamasını dokuz günlüğüne dünyadan saklamasını sağlayacaktım. Roket kalkışta patlarsa çok daha fazla zamana ihtiyacı olacağını söylemişti.

Açıklamaya geri sayım bitmişti. Moderasyonumda konuşan süper zekâ panelistleri yanımda oturuyorlardı: Eliezer Yudkowsky, Elon Musk, Nick Bostrom, Richard Mallah, Murray Shanahan, Bart Selman, Shane Legg ve Vernor Vinge. İnsanlar alkışlamayı bıraktı ama panelistler oturmayı sürdürdü çünkü onlara, sebebini açıklamadan, kalmalarını söylemiştim. Meia sonradan bana nabzının çok yükseldiğini ve masanın altında Viktoriya Krakovna'nın sakinleştirici elini yakaladığını söyledi. Bunun uğruna çalıştığımız, ümit ettiğimiz ve beklediğimiz an olduğunu biliyordum; gülümsedim.

YZ'yi faydalı tutmak için daha fazla araştırmanın gerektiğine dair toplantıda böyle bir fikir birliği olmasından ötürü ne kadar mutlu olduğumu ve hemen üzerinde çalışabileceğimiz çok fazla somut araştırma konusu olduğunu söyledim. Ancak bu oturumda ciddi riskler hakkında da konuşuldu, diye ekledim, bu yüzden de moralleri bozmamamız ve herkes bara ve dışarıda hazırladığımız yiyeceklere yönelmeden önce modumuzu yükseltmemiz gerek. "Bu yüzden de mikrofonu Elon Musk'a veriyorum!" Elon Musk mikrofonu alıp YZ güvenliği araştırmasına yüksek miktar para bağışlayacağını açıkladığında tarihin yazıldığını hissettim. Hiç de şaşırtıcı olmayan bir şekilde, çok büyük alkış aldı. Planlandığı gibi, ne kadar olduğunu söylemedi ama anlaştığımız gibi bir on milyon doların geleceğini biliyordum.

Meia ve ben konferanstan sonra İsveç ve Romanya'da ebeveynlerimizi ziyarete gittik ve babamla Stockholm'de canlı yayınlanan roket fırlatılışını nefesimizi tutarak izledik. İniş denemesi

maalesef Elon'un "hızlı ve programsız iniş" dediği şekilde bitti ve başarılı bir okyanus inişi yapmak, ekibinin on beş ayını daha aldı.[3] Ancak tüm uydular başarılı bir şekilde yörüngeye oturtulmuştu. Tıpkı bizim destek programımızın Elon'un milyonlarca takipçisiyle bir tweet yoluyla buluşması gibi.[4]



Görsel 9.2: 23 Mayıs 2014 günü, Jaan Tallinn, Anthony Aguirre, bendeniz, Meia Chita-Tegmark ve Viktoriya Krakovna FLI'nin kuruluşunu suşiyle kutluyor.

YZ Güvenliğini Gündeme Getirmek

Porto Riko Konferansı'nın ana hedefi YZ güvenliği araştırmasını ana akıma çekmekti ve bunun birkaç adımda gerçekleştiğini görmek neşe vericiydi. İlk olarak toplantının kendisi vardı. Burada pek çok araştırmacı gelişen bir camianın parçası olduklarını fark ettiklerinde konuyla haşır neşir olmak konusunda rahat hissetmeye başladılar. Birçok katılımcının desteği beni derinden etkiledi. Mesela, Cornell Üniversitesi YZ profesörü Bart Selman bana şöyle bir e-posta yolladı: "Dürüst olmak gerekirse, daha iyi

organize edilmiş ya da daha heyecan verici ve entelektüel olarak ilham verici bir bilimsel toplantıya katılmamıştım."

Bir sonraki ana akımlaştırma adımı 11 Ocak'ta, Elon, "Dünyanın en önemli yapay zekâ geliştiricileri YZ güvenliği araştırması için çağrıda bulunan açık mektubu imzalıyor," diye tweet gönderip imza sayfasının linkini verdiğinde atıldı. Bu sayfa sayesinde birçoğu dünyanın en önemli YZ üreticileri olan sekiz binin üzerinde imza geldi. Birden YZ güvenliğinden endişe eden insanların neden bahsettiklerini bilmediklerini iddia etmek zorlaştı çünkü bunu söylemek artık ileri gelen YZ araştırmacılarının neredeyse tümünün neden bahsettiklerini bilmediklerini söylemek olurdu. Açık mektup tüm dünyada medyada da yer aldı ve konferansımızdan gazetecileri uzak tuttuğumuz için mutlu olduk. Mektuptaki en panik yaratıcı kelime "gizli tehlikeler" olsa da, "Elon Musk ve Stephen Hawking Robot Ayaklanmasını Önlemek Umuduyla Açık Mektubu İmzaladılar" gibi başlıklar atıldı, katil terminatör resimleriyle birlikte servis edildi. Gördüğümüz yüzlerce makale arasından en sevdiğimiz, diğerlerini alaya alarak "iskelet gibi androidlerin insan kafataslarını ayaklarının altında ezdikleri görüntüleri canlandıran başlıklar, kompleks, dönüştürücü teknolojiyi karnaval gösterisine çevirdi," yazandı.[6] Neyse ki aklı başında makaleler de vardı ve bize başka bir zorluk getirdiler: yeni imzacılar sağanağıyla başa çıkmak. Çünkü güvenirliğimizi korumak ve "HAL 9000", "Terminatör", "Sarah Jeanette Connor" ve "Skynet" gibi muziplikleri aradan çıkarmak için her girdinin tek tek manuel olarak doğrulanması gerekiyordu. Bu ve gelecekteki açık mektuplarımız için, Viktoriya Krakovna ve Janos Kramar gönüllü bir kontrolör ekibi kurdular. Bunların arasında vardiyalı çalışan Jesse Galef, Eric Gastfriend ve Revathi Vinoth Kumar da vardı. Böylece Revathi Hindistan'da uykuya daldığında, görevi Boston'daki Eric'e devredebiliyordu.

Üçüncü ana akımlaştırma adımı dört gün sonra, Elon'un YZ güvenliği araştırmasına on milyon dolar bağış yapacağına dair açıklama linkimizi tweet attığında başladı.[7] Bir hafta sonra, dünyanın dört bir yanından araştırmacıların başvurabileceği ve bu fon için rekabet edebileceği çevrimiçi bir portal açtık. Bu başvuru sistemini hızla bir araya getirebildik çünkü Anthony ve ben benzeri yarışmaları fizik ödülleri için on yıl boyunca düzenlemiştik. Yüksek etkili hibeler üzerine odaklanmış California merkezli bir bağış kurumu olan The Open Philanthropy Project, Elon'un hediyesinin üzerine çıkmayı ve bize daha fazla hibe sağlamayı kabul etti. Kaç tane başvuru alacağımızı bilmiyorduk çünkü konu yeniydi ve zaman kısaydı. Aldığımız tepki bizi şoke etti. Yaklaşık yüz milyon dolar isteyen dünyanın dört bir yanından üç yüze yakın takım başvurmuştu. YZ profesörleri ve diğer araştırmacılardan oluşan bir panel teklifleri dikkatle inceledi ve üç yıla kadar finanse edilecek otuz yedi ekibi seçti. Kazananlar listesini ilan ettiğimizde, aktivitelerimize medyanın verdiği tepkiler ilk kez değişti ve ilk kez katil robot resimleri olmadan gerçekleşti. YZ güvenliğinin boş laf olmadığı sonunda anlaşılmaya başlıyordu: Yapılacak faydalı işler vardı ve pek çok büyük araştırma ekibi bize katılmak için kollarını sıvıyordu.

Sonraki iki yıl boyunca dördüncü ana akım adımı, bir dolu teknik yayın ve ana akım YZ konferanslarının kısımları olarak tüm dünyada YZ güvenliği üzerine düzenlenen düzinelerce atölyeyle organik biçimde gerçekleşti. İnatçı insanlar YZ camiasını güvenlik araştırmaları yapmaya yıllardır ikna etmeye çalışıyorlar ama pek başarı elde edemiyorlardı. Şimdiyse işler değişiyordu. Bu yayınların çoğu hibe programımız tarafından finanse edilmişti ve FLI'de biz de yapabildiğimiz kadar çok atölye yapmaya ya da finanse etmeye çalışıyorduk ancak giderek artan kısmı kendi vakitleri ve kaynaklarını kullanan YZ araştırmacıları tarafından

gerçekleştiriliyordu. Sonuç olarak, giderek daha çok araştırmacı kendi meslektaşlarından güvenlik araştırması hakkında bir şeyler öğreniyor, kullanışlı olmanın yanı sıra üzerine düşünülecek pek çok ilginç matematiksel ve hesaba dayalı problemler içererek eğlenceli olabileceğini keşfediyordu.

Karmaşık denklemler herkesin eğlence fikrine uymaz elbette. Porto Riko Konferansı'mızdan iki yıl sonra FLI hibesi kazananların araştırmalarını tanıtabileceği ve büyük ekranda slayt ardına slayt matematik sembollerin izlenebileceği bir teknik atölyeyle Asilomar Konferansı'mızın öncüllüğünü yaptık. Rice Üniversitesi'nde YZ profesörü olan Moşe Vardi toplantılar sıkıcılaşmaya başladığında YZ güvenliği araştırma alanını kurmayı başarmış olduğumuzu anlayacağını söyleyerek şaka yapmıştı.

YZ güvenliği çalışmalarındaki dramatik büyüme akademiyle sınırlı değildi. Amazon, DeepMind, Facebook, Google, IBM ve Microsoft faydalı YZ için bir endüstri ortaklığı başlattı.[8] Yeni, büyük YZ güvenliği bağışları, en büyük kâr amacı gütmeyen kardeş organizasyonlarımızda araştırmaların büyümesini sağladı: Berkeley'deki Machine Intelligence Research Institute, Oxford'daki Future of Humanity Institute ve Cambridge'deki (UK) Centre for the Study of Existential Risk. On milyon dolar ya da daha fazlasını içeren başka bağışlar yeni faydalı YZ çalışmalarını canlandırdı: Cambridge'deki Leverhulme Centre for the Future of Intelligence, Pittsburgh'dan K&L Gates Endowment for Ethics and Computational Technologies ve Miami'deki Ethics and Governance of Artificial Intelligence Fund. Son olarak da milyar dolarlık taahhütüyle, Elon Musk diğer girişimcilerle ortak olarak faydalı YZ'nin peşinden gidecek kâr amacı gütmeyen bir şirket olan OpenAI'ı San Francisco'da açtı. YZ güvenliği araştırması artık buradaydı.

Bu araştırma dalgasıyla beraber hem bireysel hem de kolektif biçimde açıklanan fikirler dalgası geldi. YZ üzerine sanayi

partnerliği kuruluş amaçlarını yayımladı ve tavsiye listeleriyle birlikte uzun raporlar ABD hükümeti, Stanford Üniversitesi ve IEEE (dünyanın en büyük teknik profesyoneller organizasyonu) tarafından yayımlanırken, başka yerlerden de düzinelerce rapor ve görüş yazıları basıldı.[9]

Asilomar'a katılanlar arasında anlamlı bir tartışma başlatmak ve bu kapsamlı camianın üzerine anlaştığı bir şey varsa onu öğrenmek konusunda istekliydik. Lucas Perry de bulduğumuz tüm dokümanları okuma ve fikirlerini özetleme gibi kahramanca bir işe soyundu. Anthony Aguirre tarafından başlatılan ve bir dizi uzun telekonferansla son bulan bir maratonda, FLI ekibimiz benzeri fikirleri gruplamaya ve gereksiz bürokratik zırvaları ayırarak, konuşmalarda veya başka mecralarda gayriresmî olarak ifade edilmiş etkili ama basılmamış fikirleri de dâhil edecek biçimde kısa ama öz bir ilkeler listesi oluşturmayı denediler. Ancak bu listede halen daha pek çok belirsizlik, çelişki ve yoruma açıklık mevcuttu, bu yüzden de konferanstan bir ay önce listeyi katılımcılarla paylaştık ve daha iyi ya da yenilikçi ilkeler için fikirleri ile önerilerini aldık. Bu camia katkısı, konferansta kullanmak amacıyla, önemli ölçüde revize edilmiş bir ilke listesini oluşturdu.

Asilomar'da, liste iki kademede daha iyileştirildi. İlki, detaylı düzeltmeler, geri bildirimler, yeni ilkeler ve eskilerinin rakibi olan ilkeleri üretecek şekilde, küçük gruplar, en çok ilgilendikleri ilkeler üzerine tartıştılar. İkinci olarak da tüm katılımcıların her ilkenin her versiyonuna ne derecede destek verdiklerine karar vermelerini istedik.

Bu kolektif süreç hem yorucu hem de teferruatlıydı. Ben, Anthony ve Meia, sonraki adımlar için gereken her şeyi zamanında halletmek için konferans süresince öğle yemeği ve uyku vakitlerimizden kıstık. Ama heyecan vericiydi de. Böylesine detaylı, çetrefilli ve bazen de çekişmeli tartışmalardan ve geniş bir

yelpazede geri bildirimden sonra, son soruşturmada, bir kısmının %97 destek aldığı, ilkelerin birçoğu etrafında ortaya çıkan yüksek seviye fikir birliği bizi hayrete düşürmüştü. Bu birliktelik bizi final listesine ilkeleri eklemek için yüksek bir çıta koymaya itti: Yalnızca katılımcıların en az %90'ının anlaştığı ilkeleri koruduk. Bu, benim birkaç kişisel favorim de dâhil olmak üzere, bazı popüler ilkelerin son saniyede çıkarılması anlamına gelse de toplantı salonunda dağıttığımız imza kâğıdındaki tüm ilkeleri desteklemeleri konusunda çoğu katılımcının rahat hissetmesini sağladı. İşte sonuçlar.



Görsel 9.3: Asilomar'da büyük zihinlerden oluşan gruplar YZ ilkeleri üzerine tartışıyor.

Asilomar YZ İlkeleri

Yapay zekâ dünyanın dört bir yanındaki insanların her gün kullandığı faydalı araçları çoktan sunmuştur. Bu ilkelerle sürecek gelişim önümüzdeki on yıllar ve yüzyıllarda insanlara yardımcı olacak ve onları güçlü kılacak inanılmaz fırsatlar sunacaktır.

Araştırma Sorunları

§ 1 Araştırma Hedefi: *YZ araştırmasının hedefi yönsüz zekâ değil faydalı zekâ üretmek olmalıdır.*

§ 2 Araştırma Fonlaması: *YZ'ye yapılan yatırımlar bilgisayar bilimi, ekonomi, hukuk, etik ve sosyal bilimlerin aşağıda verilen bazı çetrefilli soruları da dâhil olmak üzere, YZ'nin faydalı kullanımını sağlamak üzerine yapılacak fonlama eşliğinde olmalıdır:*

(a) *Geleceğin YZ sistemlerini arıza yapmadan ya da hacklenmeden istediğimizi yapacak şekilde yüksek oranda sağlam olarak nasıl üretebiliriz?*

(b) *İnsanların kaynaklarını ve amacını koruyarak otomasyon yoluyla zenginliğimizi nasıl artırabiliriz?*

(c) *Hukuk sistemimizi nasıl daha adil ve verimli olacak, YZ'ye ayak uyduracak ve YZ ile ilişkili riskleri yönetebilecek şekilde nasıl güncelleyebiliriz?*

(d) *YZ hangi değer setiyle hizalanmalı ve hangi hukuki ve etik statüye sahip olmalıdır?*

§ 3 Bilim-Politika Bağlantısı: *YZ araştırmacıları ve politika üreticileri arasında yapıcı ve sağlıklı bir iletişim olmalıdır.*

§ 4 Araştırma Kültürü: *YZ araştırmacıları ve geliştirmecileri arasında işbirliği, güven ve şeffaflık kültürü beslenmelidir.*

§ 5 Yarıştan Kaçmak: *YZ sistemlerini geliştiren ekipler güvenlik standartlarında kısıntıya gidilmesini engellemek için aktif biçimde işbirliği yapmalıdır.*

Etik ve Değerler

§ 6 Güvenlik: *YZ sistemleri çalışma yaşamları boyunca güvenli ve emin olmalı ve mümkün olan her koşulda bu doğrulanabilir olmalıdır.*

§ 7 Arıza Şeffaflığı: *Eğer bir YZ sistemi zarara yol açarsa, bunun sebebinin tayin edilebiliyor olması gerekir.*

§ 8 Adli Şeffaflık: *Otonom bir sistemin verebileceği her türlü adli kararlar, yetkin bir insan otorite tarafından denetlenebilecek tatmin edici bir sebep sunmalıdır.*

§ 9 Sorumluluk: *İleri YZ sistemlerinin tasarımcıları ve üreticileri kullanılmaları, suistimal edilmeleri ve eylemlerinden doğacak ahlaki sonuçlara paydaştır ve bu sonuçlara şekil verme sorumluluğu ile fırsatına sahiptir.*

§ 10 Değer Hizalama: *Yüksek oranda özerk YZ sistemleri hedefleri ve davranışlarının çalışma süreleri boyunca insan değerleriyle aynı hizada olacağından emin olacak şekilde tasarlanmalıdır.*

§ 11 İnsan Değerleri: *YZ sistemleri insanlık onuru, hakları, özgürlükleri ve kültürel çoğulculuk idealleriyle uyumlu olacak şekilde tasarlanmalı ve çalıştırılmalıdır.*

§ 12 Kişisel Mahremiyet: *YZ sistemlerinin bu veriyi analiz etme ve kullanma gücünü dikkate alarak, insanların kendi ürettikleri veriye erişme, yönetme ve kontrol etme hakkı olmalıdır.*

§ 13 Özgürlük ve Mahremiyet: *YZ'nin kişisel veriye uygulanması insanların gerçek ya da hissedilen özgürlüklerini sebepsiz yere kısıtlamamalıdır.*

§ 14 Paylaşılan Fayda: *YZ teknolojileri mümkün olduğunca çok insana fayda sağlamalı ve güçlendirmelidir.*

§ 15 Paylaşılan Zenginlik: *YZ'nin ürettiği ekonomik zenginlik tüm insanlığa fayda sağlayacak şekilde genişçe paylaşılmalıdır.*

§ 16 İnsan Kontrolü: *İnsanlar, insan seçimi amaçlara ulaşmak için kararları YZ sistemlerine devredip devretmeyeceklerini, devredeceklerse de bunu nasıl yapacaklarını seçmelidirler.*

§ 17 Yıkıcı Olmama: *İleri derecede gelişmiş YZ sistemlerinin kontrolüyle elde edilen güç saygılı olmalı, toplumun sağlığı için önem taşıyan sosyal ve medeni süreçleri altüst etmekten ziyade onları iyileştirmelidir.*

§ 18 YZ Silahlanma Yarışı: *Ölümcül otonom silahlarda bir silahlanma yarışına gitmekten kaçınılmalıdır.*

Uzun Vade Sorunlar

§ 19 Kabiliyet Tedbirleri: *Bir fikir birliği olmadığı için, gelecek YZ'nin kabiliyetlerinin üst sınırlarına dair güçlü varsayımlardan kaçınmalıyız.*

§ 20 Önem: *İleri seviye YZ Dünya'daki yaşamın tarihi üzerinde derin bir değişimi temsil edebilir ve önemiyle orantılı bir dikkat ve kaynakla planlanmalı ve yönetilmelidir.*

§ 21 Riskler: *YZ sistemlerinin ortaya çıkardığı, özellikle yıkıcı ve varoluşsal riskler beklenen etkileriyle orantılı olarak planlama ve azaltma çabalarının konusu olmalıdır.*

§ 22 Yinelemeli Öz Gelişim: *Hızla nitel ve nicel olarak artmalarına yol açabilecek biçimde yinelemeli olarak öz gelişime uğrayacak ya da kendilerini kopyalayacak şekilde tasarlanan YZ sistemleri sıkı güvenlik ve kontrol önlemlerine tabi olmalıdır.*

§ 23 Ortak Menfaat: *Süper zekâ yalnızca geniş ölçüde paylaşılan etik ideallerin hizmetinde ve tek bir devlet ya da kurum değil tüm insanlığın menfaati için geliştirilmelidir.*

İlkeleri çevrimiçi olarak paylaştıktan sonra imza listesi dramatik olarak büyüdü ve şu anda da binlerce YZ araştırmacısı ve diğer önemli düşünürleri kapsayan inanılmaz bir liste oldu. Eğer siz de imzacı olarak katılmak isterseniz, buradan yapabilirsiniz: http://futureoflife.org/ai-principles.

İlkeler konusundaki fikir birliğinin seviyesinden değil, ne kadar güçlü olduğundan da etkilenmiştik. Elbette, bazıları ilk bakışta anca, "Barış, sevgi ve annelik değerlidir," kadar ihtilaflı duruyor. Ama birçoğu da pek dişlidir, negatif versiyonlarını yap-

tığınızda da bunu daha kolay görebilirsiniz. Mesela, "Süper zekâ imkânsızdır!" madde 19'u ihlal ederken, "YZ'den gelen varoluşsal riski azaltmak için araştırma yapmak tam bir kayıptır!" madde 22'yi ihlal eder.

Gerçekten de YouTube'da uzun panel tartışmalarımızı izlerseniz göreceğiniz üzere,[11] Elon Musk, Stuart Russell, Ray Kurzweil, Demis Hassabis, Sam Harris, Nick Bostrom, David Chalmers, Bart Selman ve Jaan Tallinn, süper zekânın muhtemelen geliştirileceği ve güvenlik araştırmasının önemli olduğu konusunda hemfikirler.



Görsel 9.4: Büyüyen bir camia Asilomar'da cevapları beraber arıyor.

Dikkatli İyimserlik

Bu sonsözün başında da itiraf ettiğim gibi, uzun zamandan beri yaşamın geleceğine dair hiç bu kadar iyimser olmamıştım. Nedenini açıklamak için sizinle kişisel hikâyemi paylaştım.

Son birkaç yıldaki deneyimlerim iyimserliğimi iki ayrı sebepten ötürü artırdı. İlki, YZ camiasının yapıcı bir biçimde

önümüzdeki zorlukları aşmak için, sık sık da başka alanlardan düşünürlerle de işbirliği yaparak, inanılmaz bir biçimde bir araya geldiklerine tanıklık ettim. Elon, Asilomar toplantısından sonra YZ güvenliğinin, birkaç yıl içerisinde bir yan konu olmaktan ana akım haline gelmesine inanamadığını söyledi. Ben de en az onun kadar şaşkınım. Hem de yalnızca Bölüm 3'te bahsettiğimiz yakın dönem sorunları değil, Asilomar YZ ilkelerindeki gibi, süper zekâ ve varoluşsal riskler de saygıdeğer tartışma konuları halini alıyor. Bu ilkelerin iki yıl önce Porto Riko'da benimsenmesinin hiçbir yolu yoktu. Düşünsenize, açık mektuba girebilen en korkunç kelime bütünü "gizli tehlikelerdi."

İnsanları izlemeyi severim ve Asilomar Konferansı'nın son gününde bir ara, toplantı salonunun kenarında durdum ve YZ ve hukuk konusunda bir tartışmayı dinleyen katılımcıları izledim. Şaşırtıcı bir biçimde, sıcak ve karmaşık bir his beni esir aldı ve çok duygulandım. Porto Riko'dan çok başkaydı bu! O zaman, YZ camiasının büyük bir kesimini saygı ve korkunun bir bileşimiyle takip ettiğimi hatırlıyorum; rakip bir takım gibi değil ancak YZ'yle ilgili meslektaşlarım ve ikna etmek zorunda olduğumuz bir grup olarak. Fakat o anda hepimizin *aynı* takımda olduğu o kadar belliydi ki! Bu kitaptan öğrenmiş olacağınız üzere, muhteşem bir geleceği YZ'yle nasıl üreteceğimize dair sorunun cevabını hâlâ bilmiyorum, bu yüzden de cevapları beraber arayacağımız büyüyen bir camianın parçası olmak mutluluk verici.

Daha iyimser olmamın ikinci sebebi de FLI deneyiminin güçlendirici olmasıydı. Londra'da ağlamama sebep olan bir kaçınılmazlık hissiydi: Rahatsız edici bir geleceğin var olacağına ve yapabileceğimiz hiçbir şeyin olmadığına dair bir his. Ancak sonraki üç yıl, kaderci kasvetimi silip attı. Bir grup ücretsiz gönüllü bile zamanımızın en önemli tartışmalarından birinde pozitif bir

fark yaratabiliyorsa, hep beraber çalışırsak neler yapabileceğimizi bir düşünün!

Erik Brynjolfsson, Asilomar konuşmasında iki tür iyimserlikten bahsetti. İlki, güneşin yarın doğacağına dair olumlu beklenti gibi, koşulsuz olanı. Bir de onun "dikkatli iyimserlik" adını verdiği vardı. Bu da dikkatlı planlar ve sıkı çalışırsanız iyi şeylerin olabileceğine dair bir iyimserlikti. Yaşamın geleceğine dair hissettiğim iyimserlik işte bu tür bir iyimserlik.

YZ çağına girerken yaşamın geleceği için olumlu bir farklılık yaratmak amacıyla *siz* ne yapabilirsiniz? Birazdan açıklayacağım sebeplerden ötürü, ilk büyük adım dikkatli bir iyimser olmak için çalışmaya başlamaktır, eğer henüz değilseniz yani. Başarılı bir dikkatli iyimser olmak için, gelecek için olumlu bir vizyon geliştirmek hayati önem taşır. MIT öğrencileri kariyer tavsiyesi için ofisime geldiklerinde, genelde on yıl sonra kendilerini nerede gördüklerini sorarım ilk olarak. Eğer bir öğrenci bana, "Muhtemelen kanser koğuşunda ya da otobüs çarptıktan sonra mezarlıkta," diye cevap verecek olursa ona zor anlar yaşatırım. Yalnızca negatif bir gelecek tahayyülüne sahip olmak kariyer planlamaya korkunç bir yaklaşımdır! Tüm çabalarını, hastalıklardan ve kazalardan kaçınmak için harcamak hastalık hastalığı ve paranoya için tutulması gereken yoldur, mutluluk için değil. Bunun yerine öğrencinin hedeflerini şevkle anlatmasını isterim, işte o zaman gizli tehlikelerden kaçınarak oraya varmanın stratejilerini tartışabiliriz.

Erik, oyun teorisine göre, evlilik ve şirket birlikteliklerinden bağımsız eyaletlerin ABD'yi kurmasına dek, dünyada tüm işbirliklerinin büyük bir kısmının olumlu vizyonlardan ileri geldiğine işaret ettiğini söyledi. Sonuçta, daha büyük bir kazanç sağlanacağını düşünmüyorsanız, neden kendinizin olan bir şeyi kurban edesiniz? Bu da olumlu gelecekleri yalnızca kendimiz değil top-

lum ve insanlık için de hayal etmemiz gerektiği anlamına gelir. Diğer bir deyişle, çok daha varoluşsal umuda ihtiyacımız var! Yine de Meia'nın bana hatırlatmayı sevdiği gibi, Frankenstein'dan Terminatör'e edebiyat ve sinemadaki fütüristik vizyonlar genelde distopiktir. Diğer bir deyişle, toplum olarak geleceğimizi o farazi MIT öğrencisi kadar kötü planlıyoruz. Bu yüzden de daha çok dikkatli iyimsere ihtiyacımız var. Bu yüzden de kitap boyunca, nasıl bir gelecekten *korktuğunuzu* değil nasıl bir gelecek *istediğinizi* düşünmeniz için sizi teşvik etmeye çalıştım. Böylece, planlamak ve çalışmak için ortak hedefler bulabiliriz.

Kitapta YZ'nin bize nasıl büyük fırsatlar ve zorluklar vereceğini gördük. Temelde tüm YZ zorluklarında bize yardımcı olması muhtemelen bir strateji, YZ tamamıyla kalkışa geçmeden *önce* kendimize çekidüzen vermemiz ve insan toplumumuzu iyileştirmemizdir. Gençlerimizi, teknolojinin eline büyük bir güç vermeden önce onu sağlam ve faydalı yapmaları konusunda eğitmeliyiz. Teknoloji onları hükümsüz kılmadan önce yasalarımızı modernleştirmeliyiz. Otonom silahları içeren bir silahlanma yarışına dönüşmeden önce uluslararası çatışmaları çözmeliyiz. YZ potansiyel olarak eşitsizlikleri artırmadan önce herkes için zenginliği garanti altına alan bir ekonomi kurmalıyız. YZ güvenlik araştırma sonuçlarının görmezden gelindiği değil, uygulandığı bir toplum inşa etmeliyiz. Gelecekte, insanüstü YGZ'yle ilgili zorluklara karşı, standartlarımızı güçlü makinelere öğretmeye başlamadan önce en azından bazı temel etik standartlar konusunda anlaşmalıyız. Kutuplaşmış ve kaotik bir dünyada, YZ'yi kötü niyetleri için kullanma gücüne sahip insanlar bunu yapabilmek için daha fazla motivasyon ve yeteneğe sahip olacaklardır ve YGZ'yi inşa etmek için yarışan takımlar işbirliği yapmaktansa güvenlik konusunda kısıntıya gitme baskısını daha çok hissedeceklerdir. Özet olarak, ortak hedeflere doğru işbirliği yapan daha uyumlu

bir insan toplumu yaratabilirsek bu, YZ devriminin iyi sonuçlanma ihtimalini artıracaktır.

Diğer bir deyişle, yaşamın geleceğini iyileştirmenizin en iyi yollarından biri yarını iyileştirmektir. Bunu yapma gücüne sahipsiniz. Elbette, oy verebilir ve politikacılara eğitim, mahremiyet, ölümcül otonom silahlar, teknolojik işsizlik ve diğer konularda düşündüklerinizi söyleyebilirsiniz. Ancak her gün ne alacağınızı, hangi haberleri izleyeceğinizi, neyi paylaşacağınızı ve ne tür bir rol model olacağınızı seçerek zaten oy veriyorsunuz. Her diyaloğunu akıllı telefonunu kontrol ederek bölen biri mi olmak istersiniz yoksa teknolojiyi bilinçli ve planlı bir şekilde kullanarak kendini güçlü hisseden biri mi? Teknolojinize sahip olmak mı istersiniz yoksa teknolojinin size sahip olmasını mı? YZ çağında insan olmanın ne anlama gelmesini istersiniz? Lütfen bunu etrafınızdakilerle tartışın. Bu yalnızca önemli bir konu değil, baş döndürücü de.

Biz yaşamın geleceğinin koruyucularıyız çünkü YZ çağını biz şekillendiriyoruz. Londra'da ağlamış olsam da şu anda gelecek hakkında kaçınılmaz olan hiçbir şey olmadığını hissediyorum ve bir fark yaratmanın düşündüğümden çok daha kolay olduğunu biliyorum. Geleceğimiz taşa kazılı ve gerçekleşmek için beklemiyor, geleceği biz yaratacağız. Gelin hep beraber ilham verici bir gelecek yaratalım!

Notlar

Bölüm 1

1. Bu açık mektup, "Research Priorities for Robust and Beneficial Artificial Intelligence [Sağlam ve Faydalı Yapay Zekâ için Araştırma Öncelikleri]," şu adreste bulunabilir: http://futureoflife.org/ai-open-letter.
2. Medyadaki klasik robot yaygaracılığına örnek: Ellie Zolfagharifard, "Artificial Intelligence 'Could Be the Worst Thing to Happen to Humanity'", *Daily Mail*, 2 Mayıs 2014, http://tinyurl.com/hawkingbots.

Bölüm 2

1. YGZ teriminin kökeni üzerine: http://wp.goertzel.org/who-coined-the-term-agi.
2. Hans Moravec, "When Will Computer Hardware Match the Human Brain," *Journal of Evolution and Technology* (1998), vol. 1.
3. Hesaplama gücüne karşın yılları gösteren şekilde, 2011 yılı öncesi veri Ray Kurzweil'in kitabı *Bir Zihin Yaratmak*'tan, devamı ise şu linkteki referanslardan hesaplanmıştır: https://en.wikipedia.org/wiki/FLOPS.
4. Kuantum hesaplama öncüsü David Deutsch kuantum hesaplamanın paralel evrenler için bir kanıt olabileceğini kitabı *The Fabric of Reality: The Science of Parallel Universes—and Its Implications*'da (Londra: Allen Lane, 1997) anlatıyor. Eğer kuantum paralel evrenleri hakkında dört çoklu evren kat-

manından üçüncüsü olarak fikirlerimi öğrenmek isterseniz, bir önceki kitabıma bakabilirsiniz: Max Tegmark, *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality* (New York: Knopf, 2014).

Bölüm 3

1. YouTube'dan izleyin, "Google DeepMind's Deep Q-learning Playing Atari Breakout": https://tinyurl.com/atariai.
2. Bkz. Volodymyr Mnih et al., "Human-Level Control Through Deep Reinforcement Learning," *Nature* 518 (26 Şubat 2015): 529–533, çevrimiçi erişilebilir: http://tinyurl.com/ataripaper.
3. Big Dog robotunun hareket halindeki durumuna dair bir video: https://www.youtube.com/watch?v=W1czBcnX1Ww.
4. AlphaGo'nun inanılmaz derecede yaratıcı beşinci sütun hamlesi için, bkz. "Move 37!! Lee Sedol vs AlphaGo Match 2", https://www.youtube.com/watch?v=JNrXgpSEEIE.
5. Bilgisayarlı çeviri hakkında son gelişmeler için, bkz. Gideon Lewis-Kraus, "The Great A.I. Awakening," *New York Times Magazine* (14 Aralık 2016), çevrimiçi erişim: http://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html. GoogleTranslate'e buradan ulaşabilirsiniz: https://translate.google.com.
6. Winograd Schema Challenge yarışması: http://tinyurl.com/winogradchallenge.
7. Ariane 5 patlama videosu: https://www.youtube.com/watch?v=qnHn8W1Em6E.
8. Araştırma ekibinin Ariane 5 Uçuş 501 hata raporu: http://tinyurl.com/arianeflop.
9. NASA Mars İklim Uydusu Kaza Soruşturma Kurulu Faz I raporu: http://tinyurl.com/marsflop.
10. Mariner 1 Venüs görevi kazasına neyin yol açtığına dair en detaylı ve tutarlı açıklama, tek bir matematik sembolünün

yanlış biçimde elle transkripsiyonuydu (eksik bir üst çizgi): http://tinyurl.com/marinerflop.

11. Sovyet Phobos 1 Mars görevi kazasının detaylı bir anlatımı şurada bulunabilir: Wesley T. Huntress Jr. ve Mikhail Ya. Marov, *Soviet Robots in the Solar System* (New York: Praxis Publishing, 2011), s. 308.
12. Doğrulanmamış yazılımın Knight Capital'e kırk beş dakikada dört yüz kırk milyon dolar zarar ettirmesinin hikayesi: http://tinyurl.com/knightflop1 ve http://tinyurl.com/knightflop2.
13. Wall Street "flaş çökme" üzerine ABD hükümet raporu: "Findings Regarding the Market Events of May 6, 2010" (30 Eylül 2010), http://tinyurl.com/flashcrashreport.
14. 3D yazılan binalar (https://www.youtube.com/watch?v=SObzNdyRTBs), mikromekanik cihazlar (http://tinyurl.com/tinyprinter) ve arasındaki bir dolu başka şey (https://www.youtube.com/watch?v=xVU4FLrsPXs).
15. Toplum kökenli fab labların küresel bir haritası: https://www.fablabs.io/map.
16. Robert Williams'ın bir sanayi robotu tarafından öldürülmesi hakkında gazete haberi: http:// tinyurl.com/williamsaccident.
17. Kenji Urada'nın bir sanayi robotu tarafından öldürülmesi hakkında gazete haberi: http://tinyurl.com/uradaaccident.
18. Bir Volkswagen çalışanının bir sanayi robotu tarafından öldürülmesi hakkında gazete haberi: http://tinyurl.com/baunatalaccident.
19. İşçi ölümleri hakkında ABD hükümet raporu: https://www.osha.gov/dep/fatcat/dep_fatcat.html.
20. Araba kazası ölümlülük istatistikleri: http://tinyurl.com/roadsafety2 ve http://tinyurl.com/roadsafety3.
21. İlk Tesla pilot ölümlülüğü hakkında, bkz. Andrew Buncombe, "Tesla Crash: Driver Who Died While on Autopilot Mode 'Was Watching Harry Potter'", *Independent* (1 Temmuz 2016), http://tinyurl.com/teslacrashstory. ABD Ulusal Otoyol Trafik

Güvenliği Kurumu'nun Kusur İnceleme Birimi'nin raporu için: http://tinyurl.com/teslacrashreport.

22. *Herald of Free Enterprise* kazası için, bkz. R. B. Whittingham, *The Blame Machine: Why Human Error Causes Accidents* (Oxford, UK: Elsevier, 2004).

23 Air France 447 kazası hakkında belgesel: https://www.youtube.com/watch?v=dpPkp8OGQFI; kaza raporu: http://tinyurl.com/af447report; dışarıdan analiz: http://tinyurl.com/thomsonarticle.

24. 2003 ABD-Kanada elektrik kesintisi hakkında: http://tinyurl.com/uscanadablackout.

25. Three Mile Island kazası hakkında Başkan Komisyonu'nun final raporu: http://www.threemileisland.org/downloads/188.pdf.

26. YZ'nin, insan radyologları MRI bazlı prostat kanseri tanısında nasıl geride bırakabileceğine dair Hollanda çalışması: http://tinyurl.com/prostate-ai.

27. YZ'nin, insan patologları akciğer kanseri tanısında nasıl geride bırakabileceğine dair Stanford çalışması: http://tinyurl.com/lungcancer-ai.

28. Therac-25 radyasyon terapi kazaları soruşturması: http://tinyurl.com/theracfailure.

29. Karmaşık kullanıcı arayüzü sebebiyle ortaya çıkan Panama'daki ölümcül derecede yüksek doz radyasyon üzerine rapor: http://tinyurl.com/cobalt60accident.

30. Robotik cerrahinin yan etkileri üzerine çalışma: https://arxiv.org/abs/1507.03518.

31. Kötü hastane bakımından doğan ölüm oranları hakkında makale: http://tinyurl.com/medaccidents.

32. Yahoo bir milyar (!) kullanıcı hesabının çalındığını açıklayarak "büyük hack" konusunda yeni bir standart oluşturuyor: https://www.wired.com/2016/12/yahoo-hack-billion-users.

33. KKK katilinin beraat etmesi ve sonradan mahkûm edilmesi hakkında *New York Times* makalesi: http://tinyurl.com/kkkacquittal.
34. Aç hâkimlerin daha sert olduğuna dair Danziger et al. 2011 çalışması (http://www.pnas.org/content/108/17/6889.full), Keren Weinshall-Margel ve John Shapard (http://www.pnas.org/content/108/42/E833.full) tarafından kusurlu bulunup eleştirilmişti ancak Danziger et al. iddialarının halen geçerli olduğu konusunda ısrarcı (http://www.pnas.org/content/108/42/E834.full).
35. Pro Publica yeniden suç işleme eğilimi yazılımda ırksal önyargıya işaret ediyor: http://tinyurl.com/robojudge.
36. fMRI ve diğer beyin tarama tekniklerinin davalarda kanıt olarak kullanılması yüksek oranda tartışmalıdır, pek çok ekip %90 üzeri doğruluk payı olduğunu iddia etse de bu tekniklerin güvenilirliği de ihtilaflıdır: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00709/full.
37. PBS *The Man Who Saved the World* belgeselini tek başına Sovyet nükleer saldırısını önleyen Vasili Arkhipov hakkında yapmıştır: https://www.youtube.com/watch?v=4VPY2SgyG5w.
38. Stanislav Petrov'un ABD nükleer saldırı uyarılarını yanlış alarm olarak görmezden gelmesi hikâyesi *The Man Who Saved the World* (bir üst nottaki aynı isimli belgeselle karıştırılmamalı) isimli bir filmde anlatılır ve Petrov'a Birleşmiş Milletler Dünya Vatandaşı Ödülü vermiştir: https://www.youtube.com/watch?v=IncSjwWQHMo.
39. YZ ve robotik araştırmacılarından özerk silahlar hakkında açık mektup: http://futureoflife.org/open-letter-autonomous-weapons.
40. Bir ABD yetkilisi askerî YZ silahlanma yarışı istiyor gibi: http://tinyurl.com/workquote.
41. 1913'ten bu yana ABD'deki gelir eşitsizliği çalışması: http://gabriel-zucman.eu/files/SaezZucman2015.pdf.

42. Küresel gelir eşitsizliği üzerine Oxfam raporu: http://tinyurl.com/oxfam2017.
43. Çizilen veri şuradan alınmıştı: Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty, Emmanuel Saez ve Gabriel Zucman, *The World Wealth and Income Database* (http://www.wid.world), sermaye kazanımları da dâhildir.
44. Teknolojinin getirdiği eşitsizlik hipotezine güzel bir giriş için, bkz. Erik Brynjolfsson ve Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (New York: Norton, 2014).
45. *The Atlantic*'te çıkan az eğitimlilerin düşen maaşları hakkında makale: http://tinyurl.com/wagedrop.
46. Gelir değişiminin emekten sermayeye doğru olduğunu gösteren James Manyika sunumu: http://futureoflife.org/data/PDF/james_manyika.pdf.
47. Geleceğin iş otomasyonu hakkında Oxford Üniversitesi'nden: (http://tinyurl.com/automationoxford) ve McKinsey'den tahminler: (http://tinyurl.com/automationmckinsey).
48. Robotşef videosu: https://www.youtube.com/watch?v=fE6i2OO6Y6s.
49. Marin Soljačić bu seçenekleri 2016'de düzenlenen şu atölyede açıklamıştı: "Computers Gone Wild: Impact and Implications of Developments in Artificial Intelligence on Society" http://futureoflife.org/2016/05/06/computers-gone-wild.
50. Andrew McAfee'nin nasıl daha çok iyi iş üretebileceğimize dair teklifleri: http://futureoflife.org/data/PDF/andrew_mcafee.pdf.
51. Teknolojik işsizlik için "bu sefer farklı" diyen pek çok akademik makaleye ek olarak, "Humans Need Not Apply" kısa ve öz biçimde aynı şeyi yapar: https://www.youtube.com/watch?v=7Pq-S557XQU.
52. ABD Çalışma İstatistikleri Bürosu: http://www.bls.gov/cps/cpsaat11.htm.

53. Teknolojik işsizlik için “bu sefer farklı” argümanı: Federico Pistono, *Robots Will Steal Your Job, but That's OK* (2012), http://robotswillstealyourjob.com.
54. ABD at popülasyonunda değişiklik: http://tinyurl.com/horsedecline.
55. İşsizliğin esenliği nasıl etkilediğine dair meta analiz: Maike Luhmann et al., “Subjective Well-Being and Adaptation to Life Events: A Meta-Analysis,” *Journal of Personality and Social Psychology* 102, no. 3 (2012): 592; çevrimiçi erişim: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3289759.
56. İnsanların esenlik hissini nelerin artırdığına dair çalışmalar: Angela Duckworth, Tracy Steen ve Martin Seligman, “Positive Psychology in Clinical Practice,” *Annual Review of Clinical Psychology* 1 (2005): 629–651, çevrimiçi erişim: http://tinyurl.com/wellbeingduckworth. Weiting Ng ve Ed Diener, “What Matters to the Rich and the Poor? Subjective Well-Being, Financial Satisfaction, and Postmaterialist Needs Across the World,” *Journal of Personality and Social Psychology* 107, no. 2 (2014): 326, çevrimiçi erişim: http://psycnet.apa.org/journals/psp/107/2/326. Kirsten Weir, “More than Job Satisfaction,” *Monitor on Psychology* 44, no. 11 (Aralık 2013), çevrimiçi erişim: http://www.apa.org/monitor/2013/12/job-satisfaction.aspx.
57. Yaklaşık 10^{11} nöronu çarptığımızda, her saniye yaklaşık nöron başına 10^{4} bağlantı ve bir (10^{0}) ateşleme, insan beynini simüle etmek için yaklaşık 10^{15} FLOPS'un (1 petaFLOPS) yeterli geleceğini söyler. Ancak, ateşlemelerin detaylı zamanlaması ve nöronların ve sinapsların küçük kısımlarının da simüle edilip edilmeyeceği sorusu gibi az bilinen komplikasyonlar vardır. IBM bilgisayar bilimcisi Dharmendra Modha 38 petaFLOPS'un gerekli olacağını tahmin etmekteyken (http://tinyurl.com/javln43), sinir bilimci Henry Markram 1000 petaFLOPS'a gerek olduğunu ileri sürer (http://tinyurl.com/6rpohqv). YZ araştırmacıları Katja Grace ve Paul Christiano beyin simülasyonunun en maliyetli kısmının hesaplama değil iletişim

olacağını ve bunun da mevcut en iyi süper bilgisayarların yapabileceği bir görev olduğunu söylerler: http://aiimpacts.org/about.

58. İnsan beyninin hesaplama gücüne dair ilginç bir tahmin için: Hans Moravec, "When Will Computer Hardware Match the Human Brain," *Journal of Evolution and Technology*, vol. 1 (1998).

Bölüm 4

1. İlk mekanik kuşun videosu için, bkz. Markus Fischer, "A Robot That Flies like a Bird," TED Talk, Temmuz 2011, https://www.ted.com/talks/a_robot_that_flies_like_a_bird.

Bölüm 5

1. Ray Kurzweil, *The Singularity is Near* (New York: Viking Press, 2005).
2. Ben Goertzel'in "Nanny AI" senaryosu burada anlatılıyor: https://wiki.lesswrong.com/wiki/Nanny_AI.
3. Makineler ve insanlar arasındaki ilişki ve makinelerin bizim kölelerimiz olup olmadıkları hakkında bir tartışma için, bkz. Benjamin Wallace-Wells, "Boyhood," *New York Magazine* (20 Mayıs 2015), çevrimiçi erişim: http://tinyurl.com/aislaves.
4. Zihin suçu Nick Bostrom'un kitabı *Superintelligence*'ta ve daha teknik detaylı bir biçimde bu makalede tartışılmıştır: Nick Bostrom, Allan Dafoe ve Carrick Flynn, "Policy Desiderata for Superintelligent AI: A Vector Field Approach" (2016), http://www.nickbostrom.com/papers/aipolicy.pdf.
5. Matthew Schofield, "Memories of Stasi Color Germans' View of U.S. Surveillance Programs," *McClatchy DC Bureau* (26 Haziran 2013), çevrimiçi erişim: http://www.mcclatchydc.com/news/nation-world/national/article24750439.html.
6. İnsanların kimsenin istemediği sonuçlar üretmekte nasıl teşvik edilebileceğine dair kışkırtıcı düşünceler için, "Me-

ditations on Moloch," yazısını okumanızı öneririm: http://slatestarcodex.com/2014/07/30/meditations-on-moloch.

7. Nükleer savaşın kazaen başlayabileceği anlara dair interaktif bir zaman çizelgesi için, bkz. Future of Life Institute, "Accidental Nuclear War: A Timeline of Close Calls," çevrimiçi erişim: http://tinyurl.com/nukeoops.
8. ABD nükleer test kurbanlarına verilen tazminat ödemeleri için, bkz. ABD Adalet Bakanlığı websitesi, "Awards to Date 4/24/2015," https://www.justice.gov/civil/awards-date-04242015.
9. *Report of the Commission to Assess the Threat to the United States from Electromagnetic Pulse* (EMP) *Attack*, Nisan 2008, çevrimiçi erişim: http://www.empcommission.org/docs/A2473-EMP_Commission-7MB.pdf.
10. Hem ABD hem de Sovyet bilim insanlarının bağımsız çalışması Reagan ve Gorbaçov'u nükleer kış riski hakkında bilgilendirdi: P. J. Crutzen ve J. W. Birks, "The Atmosphere After a Nuclear War: Twilight at Noon," *Ambio* 11, no. 2/3 (1982): 114–125. R. P. Turco, O. B. Toon, T. P. Ackerman, J. B. Pollack ve C. Sagan, "Nuclear Winter: Global Consequences of Multiple Nuclear Explosions," *Science* 222 (1983): 1283–1292. V. V. Aleksandrov ve G. L. Stenchikov, "On the Modeling of the Climatic Consequences of the Nuclear War," *Proceeding on Applied Mathematics* (Moscow: Computing Centre of the USSR Academy of Sciences, 1983) 21. A. Robock, "Snow and Ice Feedbacks Prolong Effects of Nuclear Winter," *Nature* 310 (1984): 667–670.
11. Küresel bir nükleer savaşın iklim etkilerinin hesabı: A. Robock, L. Oman ve L. Stenchikov, "Nuclear Winter Revisited with a Modern Climate Model and Current Nuclear Arsenals: Still Catastrophic Consequences," *Journal of Geophysical Research* 12 (2007): D13107.

Bölüm 6

1. Daha fazla bilgi için, bkz. Anders Sandberg, "Dyson Sphere FAQ," http://www.aleph.se/Nada/dysonFAQ.html.
2. Freeman Dyson'ın, adını alan küreler hakkında yazdığı ufuk açıcı makalesi: Freeman Dyson, "Search for Artificial Stellar Sources of Infrared Radiation," *Science*, vol. 131 (1959), 1667–1668.
3. Louis Crane ve Shawn Westmoreland kara delik motorlarını açıklıyor: "Are Black Hole Starships Possible?" http://arxiv.org/pdf/0908.1803.pdf.
4. CERN'den bilinen temel parçacıkları özetleyen güzel bir infografik: http://tinyurl.com/cernparticle.
5. Nükleer bomba destekli roket itkisi fikrini anlatan nükleer olmayan Orion prototipinin eşsiz videosu: https://www.youtube.com/watch?v=E3Lxx2VAYi8.
6. Lazer yelken hakkında pedagojik bir giriş: Robert L. Forward, "Roundtrip Interstellar Travel Using Laser-Pushed Lightsails," *Journal of Spacecraft and Rockets* 21, no. 2 (Mart–Nisan 1984), çevrimiçi erişim: http://www.lunarsail.com/LightSail/rit-1.pdf.
7. Jay Olson kozmik olarak genişleyen uygarlıkları analiz ediyor: "Homogeneous Cosmology with Aggressively Expanding Civilizations," *Classical and Quantum Gravity* 32 (2015), çevrimiçi erişim: http://arxiv.org/abs/1411.4359.
8. Uzak geleceğimizin ilk derinlikli bilimsel analizi: Freeman J. Dyson, "Time Without End: Physics and Biology in an Open Universe," *Reviews of Modern Physics* 51, no. 3 (1979): 447, çevrimiçi erişim: http://blog.regehr.org/extra_files/dyson.pdf.
9. Seth Lloyd'un yukarıda bahsi geçen formülü bize hesaplama operasyonunu τ zaman aralığında yapmak, h Planck sabitiyken, $E \geq h/4\tau$ enerjiye mal olur. Eğer seri halinde T zamanında N operasyon yapmak istiyorsak, $\tau = T/N$ olur, böylece de, E/

N≥ hN/4T'dir, bu da bize E enerjisi ve T zamanı kullanarak N4 2 √ ET/h seri operasyon yapabileceğimizi söyler. Bu durumda da hem enerji hem de zaman sahibi olmanın çok fayda sağlayacağı kaynaklardır. Eğer enerjinizi *n* farklı paralel hesaplara bölerseniz, çok daha yavaş ve verimli çalışarak, N4 2 √ ET/h sonucunu verir. Nick Bostrom yüz yıllık bir insan yaşamını simüle etmenin $N=10^{27}$ operasyon gerektireceğini tahmin ediyor.

10. Eğer yaşamın kökeninin çok düşük bir ihtimal olduğuna ve en yakın komşularımızı 10^{1000} metre uzağa koyduğuna dair dikkatli bir sav görmek istiyorsanız, Princetonlı fizikçi ve astrobiyolog Edwin Turner'ın şu videosunu öneririm: "Improbable Life: An Unappealing but Plausible Scenario for Life's Origin on Earth": https://www.youtube.com/watch?v=Bt6n6Tu1beg.
11. Dünya dışı zekâ arayışı üzerine Martin Rees'in bir makalesi: https://www.edge.org/annual-question/2016/response/26665.

Bölüm 7

1. Jeremy England'ın "kayıp odaklı adaptasyon" çalışması şurada bulunabilir: Natalie Wolchover, "A New Physics Theory of Life," *Scientific American* (28 Ocak 2014), çevrimiçi erişim: https://www.scientificamerican.com/article/a-new-physics-theory-of-life. Ilya Prigogine ve Isabelle Stengers'ın *Order Out of Chaos: Man's New Dialogue with Nature* (New York: Bantam, 1984) bunun temellerini ortaya koyuyor.
2. Duygular ve fizyolojik kökenlerine dair daha fazla şey için: William James, *Principles of Psychology* (New York: Henry Holt & Co., 1890); Robert Ornstein, *Evolution of Consciousness: The Origins of the Way We Think* (New York: Simon & Schuster, 1992); António Damásio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Penguin, 2005) ve António Damásio, *Self Comes to Mind: Constructing the Conscious Brain* (New York: Vintage, 2012).

3. Eliezer Yudkowsky dost YZ'lerin hedeflerini mevcut hedeflerimizle değil *uyumlu tahminî istencimizle* (CEV) hizalamayı tartışmıştır. Basitçe söylemek gerekirse bu, daha çok şey bilseydik, daha hızlı düşünseydik ve olmak istediğimiz insanlara daha çok benzeseydik olacağımız ideal versiyonumuzun ne istediği olarak tanımlanır. Yudkowsky 2004 yılında yayımladıktan kısa bir süre sonra CEV'i eleştirmeye başlamıştır (http://intelligence.org/files/CEV.pdf), sebep olarak da hem uygulamanın zorluklarını hem de iyi tanımlanmış bir şeye yakınsama ihtimalinin belirsiz olmasını ileri sürer.
4. Ters takviyeli öğrenme yaklaşımında, temel fikir YZ'nin kendi hedeflerini gerçekleştirmeyi değil, insan sahibininkileri gerçekleştirmeyi maksimize etmeye çalışmasıdır. Bu yüzden de sahibinin isteği hakkında emin olmadığı zamanlarda dikkatli olması ve istenileni bulması için elinden geleni yapması teşvik edilir. Sahibinin onu kapatmasından da rahatsız olmamalıdır çünkü bu, sahibinin gerçekten ne istediğini yanlış anladığı anlamına gelir.
5. Steve Omohundro'nun YZ hedef oluşumu hakkındaki makalesi, "The Basic AI Drives," çevrimiçi olarak şurada bulunabilir: http://tinyurl.com/omohundro2008. Orijinal olarak şurada basılmıştır: Pei Wang, Ben Goertzel ve Stan Franklin, (ed.) *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, ed. 483–492 (Amsterdam: IOS, 2008).
6. Zekânın, etik tabanını sorgulamadan kör biçimde kurallara uymak için kullanılması durumunda ne olabileceğine dair kışkırtıcı ve tartışmalı bir kitap: Hannah Arendt, *Eichmann Kudüs'te: Kötülüğün Sıradanlığı* (İstanbul: Metis, 2018). Benzeri bir ikilem Eric Drexler'in süper zekâyı hiçbirinin büyük resmi göremeyeceği küçük parçalara ayırarak kontrol altında tutma teklifi için de geçerlidir (http://www.fhi.ox.ac.uk/reports/2015-3.pdf). Eğer bu işe yararsa, dâhilî bir ahlak pusulası olmadan, sahibinin her isteğini ahlaki bir elekten geçirmeden yapan inanılmaz güçlü bir araç olacaktır. Dis-

topik bir diktatörlükte parçalı bürokrasiyi andırır bu: Nasıl kullanılacağını bilmeden bir bölüm silah üretir, diğeri de neden mahkûm edilmiş olduğunu bilmeden mahkûmları idam eder.

7. Altın Kural'ın modern versiyonu John Rawls'ın, kim olacaklarını önceden bilmeden değiştirmeyeceği farazi bir durumun adil olacağı fikridir.
8. Mesela, Hitler'in üst düzey subaylarının çoğunun IQ'su yüksektir. Bkz. "How Accurate Were the IQ Scores of the High-Ranking Third Reich Officials Tried at Nuremberg?," *Quora*, çevrimiçi erişim: http://tinyurl.com/nurembergiq.

Bölüm 8

1. Stuart Sutherland'in bilinç üzerine yazdıkları eğlencelidir: *Macmillan Dictionary of Psychology* (London: Macmillan, 1989).
2. Kuantum mekaniğinin kurucularından Erwin Schrödinger, kitabı *Mind and Matter*'da, geçmişi düşünürken –ve bilinçli yaşam hiç var olmamış olsaydı ne olacağını tahayyül ederken- bunu demiştir. Öte yandan, YZ'nin yükselişi *gelecekte* boş sıralara oynayacak bir oyun olabileceğimiz mantıksal ihtimalini artırıyor.
3. *The Stanford Encyclopedia of Philosophy* "bilinç" kelimesinin farklı tanım ve kullanımlarına dair ayrıntılar paylaşıyor: http://tinyurl.com/stanfordconsciousness.
4. Yuval Noah Harari 2017, *Homo Deus*, s.116 (İstanbul: Kolektif, 2017)
5. Sistem 1 ve Sistem 2 üzerine bir uzmandan muazzam bir giriş: Daniel Kahneman, *Hızlı ve Yavaş Düşünme* (İstanbul, Varlık Yayınları, 2015).
6. Bkz. Christof Koch, *The Quest for Consciousness: A Neurobiological Approach* (New York: W. H. Freeman, 2004).
7. Belki de beynimize her saniye giren bilginin yalnızca küçük bir kısmından (mesela 10-50 bit) haberdarız: K. Küpfmüller,

1962, "Nachrichtenverarbeitung im Menschen," içinde *Taschenbuch der Nachrichtenverarbeitung*, K. Steinbuch (ed.), (Berlin: Springer-Verlag, 1962): 1481–1502. T. Nørretranders, *The User Illusion: Cutting Consciousness Down to Size* (New York: Viking, 1991).

8. Michio Kaku, *The Future of the Mind: The Scientific Quest to Understand, Enhance, and Empower the Mind* (New York: Doubleday, 2014); Jeff Hawkins ve Sandra Blakeslee, *On Intelligence* (New York: Times Books, 2007); Stanislas Dehaene, Michel Kerszberg ve Jean-Pierre Changeux, "A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks," *Proceedings of the National Academy of Sciences* 95 (1998): 14529–14534.

9. Penfield'ın ünlü "Yanık tostu koklayabiliyorum" deneyinin video kutlaması: https://www.youtube.com/watch?v=mSN86kphL68. Sensorimotor korteks ayrıntıları: Elaine Marieb ve Katja Hoehn, *Anatomy & Physiology*, 3rd ed. (Pearson, 2008), 391–395.

10. Bilincin sinirsel bağıntıları (NCC) çalışması son yıllarda sinirbilim camiasında ana akım hale geldi – bkz. Geraint Rees, Gabriel Kreiman ve Christof Koch, "Neural Correlates of Consciousness in Humans," *Nature Reviews Neuroscience* 3 (2002): 261–270 ve Thomas Metzinger, *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (Cambridge, MA: MIT Press, 2000).

11. Sürekli flaş baskılama nasıl çalışır: Christof Koch, *The Quest for Consciousness: A Neurobiological Approach* (W. H. Freeman, 2004); Christof Koch ve Naotsugu Tsuchiya, "Continuous Flash Suppression Reduces Negative Afterimages," *Nature Neuroscience* 8 (2005): 1096–1101.

12. Christof Koch, Marcello Massimini, Melanie Boly ve Giulio Tononi, "Neural Correlates of Consciousness: Progress and Problems," *Nature Reviews Neuroscience* 17 (2016): 307.

13. Bkz. Koch, *The Quest for Consciousness*, s. 260 ve dahası için: *The Stanford Encyclopedia of Philosophy*, http://tinyurl.com/consciousnessdelay.
14. Bilinçli algının senkronizasyonu üzerine: David Eagleman, *Beyin: Senin Hikâyen* (İstanbul: Domingo, 2016) ve *The Stanford Encyclopedia of Philosophy*, http://tinyurl.com/consciousnesssync.
15. Benjamin Libet, *Mind Time: The Temporal Factor in Consciousness* (Cambridge, MA: Harvard University Press, 2004); Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze ve John-Dylan Haynes, "Unconscious Determinants of Free Decisions in the Human Brain," *Nature Neuroscience* 11 (2008): 543–545, çevrimiçi erişim: http://www.nature.com/neuro/journal/v11/n5/full/nn.2112.html.
16. Son zamanlarda bilince yapılan teorik yaklaşım örnekleri:

    - Daniel Dennett, *Bilinç Açıklamıyor* (İstanbul: Alfa Yayınları, 2017)

    - Bernard Baars, *In the Theater of Consciousness: The Workspace of the Mind* (Oxford University Press, 2001)

    - Christof Koch, *The Quest for Consciousness: A Neurobiological Approach* (W. H. Freeman, 2004)

    - Gerald Edelman ve Giulio Tononi, *A Universe of Consciousness: How Matter Becomes Imagination* (Hachette, 2008)

    - Antonio Damasio, *Self Comes to Mind: Constructing the Conscious Brain* (New York: Vintage, 2012)

    - Stanislas Dehaene, *Bilinç ve Beyin* (İstanbul: Alfa Yayınları, 2018)

    - Stanislas Dehaene, Michel Kerszberg, ve Jean-Pierre Changeux, "A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks," *Proceedings of the National Academy of Sciences* 95 (1998) 14529–14534.

    - Stanislas Dehaene, Lucie Charles, Jean-Rémi King ve Sébastien Marti, "Toward a Computational Theory of Cons-

cious Processing," *Current Opinion in Neurobiology* 25 (2014): 760–784.

17. David Chalmers "belirme" teriminin fizik ve felsefedeki farklı kullanımlarını etraflıca tartışıyor: http://cse3521.artifice.cc/Chalmers-Emergence.pdf.
18. Bilincin çeşitli kompleks şekillerde işlenirken bilginin hissettiği şey olduğunu ileri sürüyorum: https://arxiv.org/abs/physics/0510188, https://arxiv.org/abs/0704.0646, Max Tegmark, *Our Mathematical Universe* (New York: Knopf, 2014). David Chalmers 1996 tarihli kitabı *The Conscious Mind*'da benzeri bir duyguyu ifade ediyor: "Deneyim içeriden bilgidir; fizik ise dışarıdan."
19. Adenauer Casali et al., "A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior," *Science Translational Medicine* 5 (2013): 198ra105, çevrimiçi: http://tinyurl.com/zapzip.
20. Entegre bilgi teorisi devamlı sistemler için işlemez:
    - https://arxiv.org/abs/1401.1219
    - http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00063/full
    - https://arxiv.org/abs/1601.02626
21. Kısa süreli belleği sadece otuz saniye olan Clive Wearing'le röportaj: https://www.youtube.com/watch?v=WmzU47i2xgw.
22. Scott Aaronson IIT eleştirisi: http://www.scottaaronson.com/blog/?p=1799.
23. Cerrullo IIT eleştirisi, entegrasyonun bilinç için yeterli bir koşul olmadığını savunuyor: http://tinyurl.com/cerrullocritique.
24. Simüle edilmiş insanların zombiler olacağına dair IIT tahmini: http://rstb.royalsocietypublishing.org/content/370/1668/20140167.
25. Shanahan'ın IIT eleştirisi: http://arxiv.org/ftp/arxiv/papers/1504/1504.05696.pdf.
26. Blindsight: http://tinyurl.com/blindsight-paper.

27. Belki de beynimize her saniye giren bilginin yalnızca küçük bir kısmından (mesela 10-50 bit) haberdarız: Küpfmüller, "Nachrichtenverarbeitung im Menschen"; Norretranders, *The User Illusion.*
28. "Erişim olmayan bilinci" savunan ve karşı çıkan konumlar: Victor Lamme, "How Neuroscience Will Change Our View on Consciousness," *Cognitive Neuroscience* (2010): 204–220, çevrimiçi: http://www.tandfonline.com/doi/abs/10.1080/17588921003731586.
29. "Seçici Dikkat Testi," https://www.youtube.com/watch?v=vJG698U2Mvo.
30. Bkz. Lamme, "How Neuroscience Will Change Our View on Consciousness," s.26.
31. Bu ve diğer ilgili konular Daniel Dennett's book Consciousness Explained.
32. Bkz. Kahneman, *Hızlı ve Yavaş Düşünme.*
33. *The Stanford Encyclopedia of Philosophy* özgür irade tartışmasını inceler: https://plato.stanford.edu/entries/freewill.
34. Seth Lloyd'un niçin bir YZ'nin özgür iradesi varmış gibi hissedeceğini açıkladığı video: https://www.youtube.com/watch?v=Epj3DF8jDWk.
35. Bkz. *Dreams of a Final Theory: The Search for the Fundamental Laws of Nature.*
36. Uzak geleceğimizin ilk derinlikli bilimsel analizi: Freeman J. Dyson, "Time Without End: Physics and Biology in an Open Universe," *Reviews of Modern Physics* 51, no. 3 (1979): 447, çevrimiçi erişim: http://blog.regehr.org/extra_files/dyson.pdf.

Sonsöz

1. Porto Riko Konferansı'ndan ortaya çıkan açık mektup (http://futureoflife.org/ai-open-letter) YZ sistemlerini nasıl sağlam ve faydalı yapılabileceği araştırmasının hem önemli hem de tam zamanında olduğunu ve bu öncelikler dokümanında

örnekleri verilmiş olan pek çok önemli araştırma yöneliminin mevcut olduğunu ileri sürer: http://futureoflife.org/data/documents/research_priorities.pdf.

2. Elon Musk ile YZ güvenliği üzerine yaptığım röportaj YouTube'da bulunabilir: https://www.youtube.com/watch?v=rBw0eoZTY-g.
3. Neredeyse tüm SpaceX roket iniş girişimlerinin ve ilk başarılı okyanus inişinin güzel bir video derlemesi: https://www.youtube.com/watch?v=AllaFzIPaG4.
4. Elon Musk YZ-güvenlik hibe yarışmamız hakkında tweet attı: https://twitter.com/elonmusk/status/555743387056226304.
5. Elon Musk YZ-güvenliğini destekleyen açık mektubumuz hakkında tweet attı: https://twitter.com/elonmusk/status/554320532133650432.
6. Erik Sofge "An Open Letter to Everyone Tricked into Fearing Artificial Intelligence" *Popular Science*, (14 Ocak 2015) isimli makalesinde bizim açık mektubumuz üzerinden korku yaymaya çalışanlarla alay ediyor://www.popsci.com/open-letter-everyone-tricked-fearing-ai.
7. Elon Musk Future of Life Institute ve YZ güvenlik araştırmacıları dünyasına büyük bağışı hakkında tweet attı: https://twitter.com/elonmusk/status/555743387056226304.
8. Partnership on AI'nin insanlara ve topluma faydaları üzerine daha fazla bilgi için, web sitelerine bakın: https://www.partnershiponai.org.
9. AI: One Hundred Year Study on Artificial Intelligence, Report of the 2015 Study Panel, "Artificial Intelligence and Life in 2030" (Eylül 2016) üzerine bazı güncel rapor örnekleri: http://tinyurl.com/stanfordai; YZ'nin geleceği üzerine Beyaz Saray raporu: http://tinyurl.com/obamaAIreport; YZ ve işler üzerine Beyaz Saray raporu: http://tinyurl.com/AIjobsreport; YZ ve insan esenliği üzerine IEEE raporu, "Ethically Aligned Design, Version 1" (13 Aralık 2016): http://standards.

ieee.org/develop/indconn/ec/ead_v1.pdf; US Robotics için yol haritası: http://tinyurl.com/roboticsmap.

10. Son elemeyi geçemeyen ilkeler arasında, favorilerimden biri şuydu: "Bilinç tedbiri: Hiçbir fikir birliği olmazsa, ileri YZ'nin bilinç ya da hislere sahip olup olmadığı ya da bunlara gereksinim duyup duymadığı üzerine güçlü varsayımlardan kaçınmalıyız." Pek çok değişikliğe uğradı ama sonuncusunda, ihtilaflı kelime "bilincin" yerini "öznel deneyim" aldı – ama bu ilke yine de %88'de kaldı, %90 barajını kılpayı kaçırdı.
11. Süper zekâ üzerine tartışma panelinde Elon Musk ve diğer büyük beyinler vardı: http://tinyurl.com/asilomarAI.

İndeks

C

F

I

İ

J

K

L

M

N

O

Ö

P

Q

R

S

Ş

T

U

Ü

V

W

Y

Z



MAX TEGMARK MIT'de fizik profesörüdür ve Yaşamın Geleceği Enstitüsü'nün başkanıdır. *Our Mathematical Universe* kitabının yazarıdır ve düzinelerce bilimsel belgeselde yer almıştır. Fikirler, macera ve gelecek hakkındaki tutkusu bulaşıcıdır.

Kapak görselleri: Suvadip Das; (adam) Netfalls Remy Musser/Shutterstock

Gelecek Yapay Zekâ'da fakat bu gelecek nasıl olacak?

Süper zekâ bizim kölemiz mi yoksa efendimiz mi olacak?

Faydalı YZ araştırmalarının gündeme oturmasını sağlayan MIT Profesörü Max Tegmark, bizi YZ üzerine var olan son tartışmaların merkezine götürüyor. Varlığımızın gelecekteki safhasını keşfedebilmek için efsaneleri gerçeklerden, ütopyaları distopyalardan ayırıyor. Yapay Zekâ suç, savaş, adalet, iş gücü, toplum ve bizim insan olma hissimizi nasıl etkileyecek? YZ'nin yükselişi diğer tüm teknolojilerden daha çok geleceğimizi şekillendirme potansiyeline sahip.

Otomasyon sayesinde refah seviyemizi artırırken insanların işsiz ve amaçsız kalmasını nasıl önleyebiliriz? Çocuklarımıza nasıl kariyer tavsiyeleri vermeliyiz? Çökmeden, hata yapmadan veya hacklenmeden bizim istediklerimizi yapabilmeleri için gelecekteki YZ sistemlerini nasıl daha fazla güçlendirebiliriz? Ölümcül otonom silah sistemlerinde bir silahlanma yarışından korkmalı mıyız? Makineler zamanla her alanda bizi geçip işlerimizi, hatta gezegendeki yerimizi elimizden alacak mı? YZ daha önce eşi görülmemiş bir şekilde yaşamın gelişmesine yardım mı edecek yoksa bize yönetebileceğimizden daha büyük bir güç mü verecek?

Nasıl bir gelecek istiyorsunuz? Bu kitap, çağımızın belki de en önemli tartışmasına katılmanıza imkân veriyor. Tüm bakış açılarına yer verirken, süper zekâ, anlam, bilinç ve evrendeki yaşamın fiziksel sınırları gibi en radikal konulara değinmekten kaçınmıyor.

"Bu kitap, Dünya ve ötesinde yaşamın, zekânın ve bilincin geleceği maceramızda karşılaşabileceğimiz güçlükler ve seçimler üzerine muhteşem bir rehber."
Elon Musk, SpaceX ve Tesla'nın CEO'su

"Sadece bilim insanları, sanayiciler ve generaller değil, hepimiz kendimize gelecekteki Yapay Zekâ'nın risklerinden kaçınmak ve faydalarından yararlanabilmek için neler yapabiliriz diye sormalıyız. Bu, ç[illegible]n önemli tartışması ve Tegmark'ın kitabı bu tartışmaya katılabilmenize yardım edecek."
Prof. Stephen Hawking



www.pegasusyayinlari.com

